

OTHER INFORMATION: promoter- Arabidopsis agamous  
OTHER INFORMATION: (GenBank ATGAMSG)

NAME/KEY: Other

LOCATION: 448...2013

OTHER INFORMATION: coding sequence- Arabidopsis  
OTHER INFORMATION: apetalas3 (GenBank ATHPTALA)

NAME/KEY: Other

LOCATION: 2020...2286

OTHER INFORMATION: No. 6140554 terminator

US-09-049-475-5

Query Match

Best Local Similarity 1.8%; Score 25; DB 3; Length 2319;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAAATAT 1251

DB 1395 CACACACACACACACACAAATAT 1371

RESULT 2

US-09-734-675-3/C

Sequence 3, Application US/09734675

Patent No. 6365391

GENERAL INFORMATION:

APPLICANT: WEBSTER, Marion et al

TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,

TITLE OF INVENTION: NUCLEIC/ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND

FILE REFERENCE: C1000862

CURRENT APPLICATION NUMBER: US/09/734,675

CURRENT FILING DATE: 2000-12-13

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 3

LENGTH: 38844

TYPE: DNA

ORGANISM: Human

US-09-734-675-3

Query Match

Best Local Similarity 1.8%; Score 25; DB 4; Length 38844;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAAATAT 1251

DB 16438 CACACACACACACACACAAATAT 16414

RESULT 3

US-09-009-913-1/C

Sequence 1, Application US/09009913

Patent No. 6087485

GENERAL INFORMATION:

APPLICANT: AXYIS Pharmaceuticals, Inc.

TITLE OF INVENTION: Asthma Related Genes

NUMBER OF SEQUENCES: 339

CORRESPONDENCE ADDRESS:

ADDRESSEE: Bozicevic & Reed, LLP

STREET: 285 Hamilton Ave, Suite 200

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/009,913

FILING DATE: 21-JAN-1998

CLASSIFICATION:

*Filing date Jan 21, 1998  
102 (e)*

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sherwood, Pamela J

REGISTRATION NUMBER: 36,677

REFERENCE/DOCKET NUMBER: SEQ-4P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-327-3231

TELEFAX: 650-327-3231

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 72928 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

US-09-009-913-1

Query Match

Best Local Similarity 1.8%; Score 25; DB 3; Length 72928;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAAATAT 1251

DB 159 CACACACACACACACACAAATAT 135

RESULT 4

US-08-118-200-12/C

Sequence 12, Application US/08118200

Patent No. 6197500

GENERAL INFORMATION:

APPLICANT: SUTHERLAND, Grant R

APPLICANT: RICHARDS, Robert J

APPLICANT: SCHLESSINGER, David

APPLICANT: NAGARAJA, Ramalah

APPLICANT: KREMER, Eric J

APPLICANT: YO, Sul

APPLICANT: BAKER, Elizabeth

APPLICANT: MULLEY, John C

APPLICANT: MANDEL, Jean-Louis

APPLICANT: PRITCHARD, Melanie April

APPLICANT: LYNCH, Michael

TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED

TITLE OF INVENTION: FRAGILE X SYNDROME

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,200

FILING DATE: 09-SEP-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/802,650

FILING DATE: 05-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/672,232

FILING DATE: 20-MAR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/638,518

FILING DATE: 04-JAN-1991

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic acid search, using sw model

Run on: January 18, 2003, 21:47:29 ; Search time 78 Seconds  
(without alignments)  
5433.683 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382  
Sequence: 1 catatgctaaagcctaaa.....taagcaatgaattcctcg 1382

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Hit size: 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents.NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1.8	2319	US-09-049-475-5	Sequence 5, Appl1
2	1.8	38844	US-09-734-675-3	Sequence 3, Appl1
3	1.8	72928	US-09-009-913-1	Sequence 1, Appl1
4	1.7	360	US-08-118-200-12	Sequence 12, Appl1
5	1.7	360	US-08-458-745-12	Sequence 12, Appl1
6	1.7	1611	US-08-650-598-1	Sequence 1, Appl1
7	1.7	6550	US-09-097-319A-1	Sequence 1, Appl1
8	1.7	225	US-08-232-117A-27	Sequence 27, Appl1
9	1.7	240	US-08-232-177A-32	Sequence 32, Appl1
10	1.7	487	US-08-599-252-108	Sequence 108, App
11	1.7	487	PCT-US96-06352-108	Sequence 108, App
12	1.7	487	PCT-US96-06583-108	Sequence 108, App
13	1.7	672	US-08-599-252-102	Sequence 102, App
14	1.7	672	PCT-US96-06352-102	Sequence 102, App
15	1.7	672	PCT-US96-06583-102	Sequence 102, App
16	1.7	1166	US-08-599-252-95	Sequence 95, Appl1
17	1.7	1166	PCT-US96-06352-95	Sequence 95, Appl1
18	1.7	1166	PCT-US96-06583-95	Sequence 95, Appl1
19	1.7	1212	US-09-218-467B-5	Sequence 5, Appl1
20	1.7	1647	US-08-198-446B-9	Sequence 9, Appl1
21	1.7	1647	US-08-870-693-9	Sequence 9, Appl1
22	1.7	2040	US-08-393-983-17	Sequence 17, Appl1
23	1.7	2529	US-08-883-795A-37	Sequence 37, Appl1
24	1.7	2967	US-08-637-823B-26	Sequence 26, Appl1
25	1.7	3288	US-08-620-694A-1	Sequence 1, Appl1
26	1.7	3288	US-09-022-255-1	Sequence 1, Appl1
27	1.7	3288	US-09-022-696-1	Sequence 1, Appl1

28	23	1.7	3288	3	US-08-978-773-1	Sequence 1, Appl1
29	23	1.7	3288	3	US-09-022-253-1	Sequence 1, Appl1
30	23	1.7	3288	3	US-09-022-260-1	Sequence 1, Appl1
31	23	1.7	3288	4	US-09-022-259-1	Sequence 1, Appl1
32	23	1.7	3288	4	US-09-022-257-1	Sequence 1, Appl1
33	23	1.7	3381	4	US-09-009-119-1	Sequence 1, Appl1
34	23	1.7	3381	4	US-09-371-507-1	Sequence 1, Appl1
35	23	1.7	3383	5	PCT-US95-09098-1	Sequence 1, Appl1
36	23	1.7	3468	4	US-09-218-467B-1	Sequence 1, Appl1
37	23	1.7	3740	4	US-09-162-274A-6	Sequence 6, Appl1
38	23	1.7	4771	2	US-08-866-650-2	Sequence 2, Appl1
39	23	1.7	4771	2	US-09-021-287-2	Sequence 2, Appl1
40	23	1.7	4771	4	US-09-240-473-2	Sequence 2, Appl1
41	23	1.7	5578	1	US-08-081-610-2	Sequence 2, Appl1
42	23	1.7	7970	1	US-08-135-511-31	Sequence 31, Appl1
43	23	1.7	7997	1	US-08-187-453-31	Sequence 31, Appl1
44	23	1.7	7997	1	US-08-562-985A-7	Sequence 7, Appl1
45	23	1.7	16442	3	US-08-781-891-208	Sequence 208, App

#### ALIGNMENTS

RESULT 1  
US-09-049-475-5/C  
Sequence 5, Application US/09049475  
Patent No. 6140534  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, David  
TITLE OF INVENTION: IMPROVEMENTS RELATING TO THE  
TITLE OF INVENTION: SPECIFICITY OF GENE EXPRESSION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS: PENNIE & EDMONDS LLP  
ADDRESS: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/049,475  
FILING DATE: 27-MAR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9706381.2  
FILING DATE: 27-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 9341-014-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2319 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: Other  
ORIGINAL SOURCE:  
ORGANISM: Arabidopsis thaliana  
STRAIN: Landsberg  
FEATURE:  
NAME/KEY: Other  
LOCATION: 26...441

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,177A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/341,562
FILING DATE: 21-APR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REGISTRATION NUMBER: 30,492
REFERENCE/DOCKET NUMBER: 09865,601
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 831-2100
TELEFAX: (608) 831-2106
TELEX:
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Caucasian
TISSUE TYPE: Blood
IMMEDIATE SOURCE:
CLONE: Mtd49
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 15
FEATURE:
NAME/KEY: repeat_region
LOCATION: 82..125
OTHER INFORMATION: /rpt_type="tandem"
OTHER INFORMATION: /rpt_family="(dc-da)n.(dg-dt)n"
OTHER INFORMATION: /citation="(12)"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 61..81
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="PCR primer"
OTHER INFORMATION: /citation="(11)"
FEATURE:
NAME/KEY: misc_feature
LOCATION: complement(128..147)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="PCR primer"
OTHER INFORMATION: /citation="(11)"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..240
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /standard_name="Only one strand sequenced"
PUBLICATION INFORMATION:
AUTHORS: Weber, J. L.
AUTHORS: Kwitek, A. E.
AUTHORS: May, P. E.
TITLE: Dinucleotide repeat polymorphism at the
JOURNAL: Nucleic Acids Res.
VOLUME: 18
PAGES: 4640-
DATE: 1990
PUBLICATION INFORMATION:
AUTHORS: Weber, James L.
AUTHORS: May, Paula E.
TITLE: Abundant Class of Human DNA Polymorphisms

```

TITLE: Which Can Be Typed Using the Polymerase Chain  
JOURNAL: Am. J. Hum. Genet.  
VOLUME: 44  
PAGES: 398-396  
DATE: 1989  
US-08-222-177A-32

Query Match 1.7%; Score 23; DB 1; Length 240;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAAT 1249  
DB 106 CACACACACACACACACAAAT 128

RESULT 10  
US-08-599-252-108

Sequence 108, Application US/08599252  
Patent No. 5705343

## GENERAL INFORMATION:

APPLICANT: DRAVNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GNIKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
TELEPHONE: (202) 887-0763  
TELEFAX: (202) 887-0763  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/599,252  
FILING DATE: 09-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 487 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-599-252-108

Query Match 1.7%; Score 23; DB 1; Length 487;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAAT 1249  
DB 322 CACACACACACACACACAAAT 344

RESULT 11

PCT-US96-06352-108

Sequence 108, Application PC/TUS9606352  
GENERAL INFORMATION:

APPLICANT: DRAVNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GNIKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06352  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,252  
FILING DATE: 09-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 487 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US96-06352-108

Query Match 1.7%; Score 23; DB 5; Length 487;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAAT 1249  
DB 322 CACACACACACACACACAAAT 344

RESULT 12  
PCT-US96-06583-108

Sequence 108, Application PC/TUS9606583  
GENERAL INFORMATION:

APPLICANT: DRAVNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GNIKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 21:47:29 : Search time 78 seconds  
(without alignments)  
5433.683 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382

Sequence: 1 catatgtctaagacataa.....taagcaatlaattctcctg 1382

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCrUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	1.8	2319	3	US-09-049-475-5
C 2	25	1.8	38844	4	US-09-734-675-3
C 3	25	1.8	72928	3	US-09-009-913-1
C 4	24	1.7	360	4	US-08-118-200-12
C 5	24	1.7	360	4	US-08-458-745-12
C 6	24	1.7	1611	2	US-08-650-598-1
C 7	24	1.7	6550	4	US-09-097-319A-1
8	23	1.7	225	1	US-08-222-177A-27
9	23	1.7	487	1	US-08-222-177A-32
10	23	1.7	487	1	US-08-599-252-108
11	23	1.7	487	5	PCT-US86-06352-108
12	23	1.7	487	5	PCT-US86-06353-108
13	23	1.7	672	23	PCT-US96-06352-102
14	23	1.7	672	5	PCT-US96-06352-102
15	23	1.7	672	5	PCT-US96-06353-102
16	23	1.7	1166	1	US-08-599-252-95
17	23	1.7	1166	5	PCT-US86-06352-95
18	23	1.7	1166	5	PCT-US86-06353-95
19	23	1.7	1212	4	US-09-218-467B-5
20	23	1.7	1647	1	US-08-198-446B-9
21	23	1.7	1647	2	US-08-870-693-9
22	23	1.7	2040	1	US-08-393-985-17
23	23	1.7	2529	2	US-08-883-795A-37
24	23	1.7	2567	4	US-08-637-823B-26
25	23	1.7	3288	2	US-08-620-694A-1
26	23	1.7	3288	3	US-09-022-255-1
27	23	1.7	3288	3	US-09-022-696-1

C 28	23	1.7	3288	3	US-08-978-773-1	Sequence 1, Appl1
C 29	23	1.7	3288	3	US-09-022-253-1	Sequence 1, Appl1
C 30	23	1.7	3288	3	US-09-022-260-1	Sequence 1, Appl1
C 31	23	1.7	3288	4	US-09-022-259-1	Sequence 1, Appl1
C 32	23	1.7	3288	4	US-09-022-257-1	Sequence 1, Appl1
C 33	23	1.7	3381	4	US-09-009-119-1	Sequence 1, Appl1
C 34	23	1.7	3381	4	US-09-371-507-1	Sequence 1, Appl1
C 35	23	1.7	3383	5	PCT-US95-09098-1	Sequence 1, Appl1
36	23	1.7	3488	4	US-09-218-467B-1	Sequence 1, Appl1
37	23	1.7	3740	4	US-09-162-274A-6	Sequence 6, Appl1
C 38	23	1.7	4771	2	US-08-866-650-2	Sequence 2, Appl1
C 39	23	1.7	4771	2	US-09-021-287-2	Sequence 2, Appl1
C 40	23	1.7	4771	4	US-09-240-473-2	Sequence 2, Appl1
C 41	23	1.7	5578	1	US-08-081-610-2	Sequence 2, Appl1
C 42	23	1.7	7870	1	US-08-135-511-31	Sequence 31, Appl1
C 43	23	1.7	7997	1	US-08-187-453-31	Sequence 31, Appl1
C 44	23	1.7	7997	1	US-08-562-985A-7	Sequence 7, Appl1
C 45	23	1.7	16442	3	US-08-781-891-208	Sequence 208, App

## ALIGNMENTS

RESULT 1  
US-09-049-475-5/C  
Sequence 5, Application US/09049475  
Patent No. 6140534  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, David  
APPLICANT: Thomas, Christopher  
TITLE OF INVENTION: IMPROVEMENTS RELATING TO THE  
TITLE OF INVENTION: SPECIFICITY OF GENE EXPRESSION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/049,475  
FILING DATE: 27-MAR-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9706381.2  
FILING DATE: 27-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
TELEPHONE/DOCKET NUMBER: 9341-014-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2319 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: both  
MOLECULE TYPE: Other  
ORGANISM: Arabidopsis thaliana  
STRAIN: Landsberg  
FEATURE:  
NAME/KEY: Other  
LOCATION: 26...441

OTHER INFORMATION: promoter- Arabidopsis agamous  
OTHER INFORMATION: (GenBank ATAGAMSC)  
NAME/KEY: other  
LOCATION: 448...2013  
OTHER INFORMATION: coding sequence- Arabidopsis  
OTHER INFORMATION: apetalas3 (GenBank ATPETALA)  
NAME/KEY: other  
LOCATION: 2020...2286  
OTHER INFORMATION: No. 6140554 terminator  
US-09-049-475-5

Query Match 1.8%; Score 25; DB 3; Length 2319;  
Best Local Similarity 100.0%; Pred. No. 0.0059;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1227 CACACACACACACACACAATAT 1251  
|||||  
DB 1395 CACACACACACACACACAATAT 1371

RESULT 2  
US-09-734-675-3/c  
Sequence 3, Application US/09734675  
Patent No. 6363591  
GENERAL INFORMATION:  
APPLICANT: WEBSTER, Marion et al  
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
TITLE OF INVENTION: USES THEREOF  
FILE REFERENCE: C1000862  
CURRENT APPLICATION NUMBER: US/09/734,675  
CURRENT FILING DATE: 2000-12-13  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 38844  
TYPE: DNA  
ORGANISM: Human  
US-09-734-675-3

Query Match 1.8%; Score 25; DB 4; Length 38844;  
Best Local Similarity 100.0%; Pred. No. 0.0063;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1227 CACACACACACACACACAATAT 1251  
|||||  
DB 16438 CACACACACACACACACAATAT 16414

RESULT 3  
US-09-009-913-1/c  
Sequence 1, Application US/09009913  
Patent No. 6087485  
GENERAL INFORMATION:  
APPLICANT: Axxs Pharmaceuticals, Inc.  
TITLE OF INVENTION: Asthma Related Genes  
NUMBER OF SEQUENCES: 339  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bozicevic & Reed, LLP  
STREET: 285 Hamilton Ave, Suite 200  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/009,913  
FILING DATE: 21-JAN-1998  
CLASSIFICATION:

*Filing date Jan 21, 1998  
102ce)*

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: SEQ-4P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-327-3231  
TELEFAX: 650-327-3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 72928 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-09-009-913-1

Query Match 1.8%; Score 25; DB 3; Length 72928;  
Best Local Similarity 100.0%; Pred. No. 0.0064;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1227 CACACACACACACACACAATAT 1251  
|||||  
DB 159 CACACACACACACACACAATAT 135

RESULT 4  
US-08-118-200-12/c  
Sequence 12, Application US/08118200  
Patent No. 6197500  
GENERAL INFORMATION:  
APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I  
APPLICANT: SCHLESSINGER, David  
APPLICANT: NAGARAJA, Ramalath  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
TITLE OF INVENTION: FRAGILE X SYNDROME  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,200  
FILING DATE: 09-SEP-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-118-200-12

Query Match 1.7%; Score 24; DB 4; Length 360;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAAATA 1250  
|||||  
DB 269 CACACACACACACACACAAATA 246

RESULT 5  
US-08-458-745-12/c  
Sequence 12, Application US/08458745  
Patent No. 6242576  
GENERAL INFORMATION:  
APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I  
APPLICANT: SCHLESINGER, David  
APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
TITLE OF INVENTION: FRAGILE X SYNDROME  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,745  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/118,200  
FILING DATE: 09-SEP-1993  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518

FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 360 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-745-12

Query Match 1.7%; Score 24; DB 4; Length 360;  
Best Local Similarity 100.0%; Pred. No. 0.018;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAAATA 1250  
|||||  
DB 269 CACACACACACACACACAAATA 246

RESULT 6  
US-08-650-598-1/c  
Sequence 1, Application US/08650598  
Patent No. 5877020  
GENERAL INFORMATION:  
APPLICANT: Allitalo, Karl  
TITLE OF INVENTION: Promoter of the Receptor Tyrosine Kinase, TIE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/650,598  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/310,717  
FILING DATE: 22-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gass, David A.  
REGISTRATION NUMBER: 38,153  
REFERENCE/DOCKET NUMBER: 28113/33245  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1611 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-650-598-1

Query Match 1.7%; Score 24; DB 2; Length 1611;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACCAATA 1250  
DB 1095 CACACACACACACCAATA 1072

## RESULT 7

US-09-097-319A-1/C  
; Sequence 1, Application US/09097319A  
; Patent No. 6384207  
; GENERAL INFORMATION:  
; APPLICANT: Ainley, Michael  
; APPLICANT: Armstrong, Katherine  
; APPLICANT: Belmar, Scott  
; APPLICANT: Folkerts, Otto  
; APPLICANT: Hopkins, Nicole  
; APPLICANT: Menke, Michael A.  
; APPLICANT: Paredy, Dayakar  
; APPLICANT: Petolino, Joseph F.  
; APPLICANT: Smith, Kelley  
; APPLICANT: Woosley, Aaron  
; TITLE OF INVENTION: Regulatory Sequences for Transgenic Plants  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Dowlanco Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/097,319A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R.  
; TELEPHONE: 317 337 4816  
; TELEFAX: 317 337 4847  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6550 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 4201..4425  
; OTHER INFORMATION: /product= "peroxidase"  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 4426..5058  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 5059..5250  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 5251..5382  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 5383..5548  
; FEATURE:  
; NAME/KEY: intron  
; LOCATION: 5549..5649

FEATURE:  
; NAME/KEY: exon  
; LOCATION: 5650..6065  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(4201..4425, 5059..5250, 5383..5547, 5649  
; ; LOCATION: ..6068)  
; US-09-097-319A-1

Query Match 1.7%; Score 24; DB 4; Length 6550;  
Best Local Similarity 100.0%; Pred. No. 0.019;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1228 ACACACACACACCAATAT 1251  
DB 5322 ACACACACACACCAATAT 5299

## RESULT 8

US-08-222-177A-27  
; Sequence 27, Application US/08222177A  
; Patent No. 5582979  
; GENERAL INFORMATION:  
; APPLICANT: Weber, James L.  
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN  
; ; TITLE OF INVENTION: (dc-da)n.(dg-dt)n SEQUENCES AND METHODS OF USING SAME  
; NUMBER OF SEQUENCES: 460  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Demilt Ross & Stevens, S.C.  
; STREET: 8000 Excelsior Drive, Suite 401  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53717-1914  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/222,177A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/341,562  
; FILING DATE: 21-Apr-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sara, Charles S.  
; REGISTRATION NUMBER: 30,492  
; REFERENCE/DOCKET NUMBER: 09865.601  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 831-2100  
; TELEFAX: (608) 831-2106  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; INDIVIDUAL ISOLATE: Caucasian  
; TISSUE TYPE: Blood  
; IMMEDIATE SOURCE:  
; CLONE: Mtd43  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: 5q  
; FEATURE:  
; NAME/KEY: repeat\_region  
; LOCATION: 128..160  
; OTHER INFORMATION: /rpt\_type= "tandem"

```
;
; OTHER INFORMATION: /rpl_family="(dc-da)n.(dg-dt)n"
; OTHER INFORMATION: /citation="(12)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 43..62
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="PCR primer"
; OTHER INFORMATION: /citation="(11)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: complement(160..181)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="PCR primer"
; OTHER INFORMATION: /citation="(11)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..225
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="Only one strand sequenced"
; PUBLICATION INFORMATION:
; AUTHORS: Weber, J. L.
; AUTHORS: Polymeropoulos, M. H.
; AUTHORS: May, P. E.
; AUTHORS: Kwik, A. E.
; AUTHORS: Xiao, H.
; AUTHORS: McPherson, J. D.
; AUTHORS: Wasmuth, J. J.
; TITLE: Mapping of human chromosome 5 microsatellite
; TITLE: polymorphisms
; JOURNAL: Genomics
; DATE: 1991
; PUBLICATION INFORMATION:
; AUTHORS: Weber, James L.
; AUTHORS: May, Paula E.
; TITLE: Abundant Class of Human DNA Polymorphisms
; TITLE: Which Can Be Typed Using the Polymerase Chain
; TITLE: Reaction
; JOURNAL: Am. J. Hum. Genet.
; VOLUME: 44
; PAGES: 388-396
; DATE: 1989
; US-08-222-177A-27
;
; Query Match 1.7%; Score 23; DB 1; Length 225;
; Best Local Similarity 100.0%; Pred. No. 0.057;
; Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1224 ACGCACACACACACACACACA 1246
; DB 125 ACGCACACACACACACACACA 147
;
; RESULT 9
; US-08-222-177A-32
; Sequence 33, Application US/08222177A Dec 10, 1996 10266)
; Patent No. 5582979
; GENERAL INFORMATION:
; APPLICANT: Weber, James L.
; TITLE OF INVENTION: LENGTH POLYMORPHISMS IN
; TITLE OF INVENTION: (dc-da)n.(dg-dt)n SEQUENCES AND METHODS OF USING SAME
; NUMBER OF SEQUENCES: 460
; CORRESPONDENCE ADDRESS:
; ADDRESS: Demilt Ross & Stevens, S.C.
; STREET: 8000 Excelstor Drive, Suite 401
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,177A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/341,562
; FILING DATE: 21-APR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REGISTRATION NUMBER: 30,492
; REFERENCE/DOCKET NUMBER: 09865.601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 831-2100
; TELEFAX: (608) 831-2106
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Caucasian
; TISSUE TYPE: Blood
; IMMEDIATE SOURCE:
; CLONE: MEd49
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 15
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 82..125
; OTHER INFORMATION: /rpl_type="random"
; OTHER INFORMATION: /rpl_family="(dc-da)n.(dg-dt)n"
; OTHER INFORMATION: /citation="(12)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 61..81
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="PCR primer"
; OTHER INFORMATION: /citation="(11)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: complement(128..147)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="PCR primer"
; OTHER INFORMATION: /citation="(11)"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..240
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /standard_name="Only one strand sequenced"
; PUBLICATION INFORMATION:
; AUTHORS: Weber, J. L.
; AUTHORS: Kwik, A. E.
; AUTHORS: May, P. E.
; TITLE: Dinucleotide repeat polymorphism at the
; JOURNAL: Nucleic Acids Res.
; VOLUME: 18
; PAGES: 4640-
; DATE: 1990
; PUBLICATION INFORMATION:
; AUTHORS: Weber, James L.
; AUTHORS: May, Paula E.
; TITLE: Abundant Class of Human DNA Polymorphisms
```

;; TITLE: Which Can Be Typed Using the Polymerase Chain  
;; JOURNAL: Am. J. Hum. Genet.  
;; VOLUME: 44  
;; PAGES: 388-396  
;; DATE: 1989  
US-08-222-177A-32

Query Match 1.7%; Score 23; DB 1; Length 240;  
Best Local Similarity 100.0%; Pred. No. 0.057;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1227 CACACACACACACACAAAT 1249  
|||||  
DB 106 CACACACACACACACAAAT 128

RESULT 10  
US-08-599-252-108  
; Sequence 108, Application US/08599252  
; Patent No. 5705343  
; GENERAL INFORMATION:  
; APPLICANT: DRAYNA, DENNIS T.  
; APPLICANT: FEDER, JOHN N.  
; APPLICANT: GNIKE, ANDREAS  
; APPLICANT: KIMMEL, BRUCE E.  
; APPLICANT: THOMAS, WINSTON J.  
; APPLICANT: WOLFE, ROGER K.  
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
; TITLE OF INVENTION: HEMOCHROMATOSIS  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,252  
; FILING DATE: 09-FEB-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 9053-0001.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 108:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-599-252-108

Query Match 1.7%; Score 23; DB 1; Length 487;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1227 CACACACACACACACAAAT 1249  
|||||  
DB 322 CACACACACACACACAAAT 344

RESULT 11

PCT-US96-06352-108  
; Sequence 108, Application PC/TUS9606352  
; GENERAL INFORMATION:  
; APPLICANT: DRAYNA, DENNIS T.  
; APPLICANT: FEDER, JOHN N.  
; APPLICANT: GNIKE, ANDREAS  
; APPLICANT: KIMMEL, BRUCE E.  
; APPLICANT: THOMAS, WINSTON J.  
; APPLICANT: WOLFE, ROGER K.  
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
; TITLE OF INVENTION: HEMOCHROMATOSIS  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/06352  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,252  
; FILING DATE: 09-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 9053-0001.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 108:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
PCT-US96-06352-108

Query Match 1.7%; Score 23; DB 5; Length 487;  
Best Local Similarity 100.0%; Pred. No. 0.058;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1227 CACACACACACACACAAAT 1249  
|||||  
DB 322 CACACACACACACACAAAT 344

RESULT 12  
PCT-US96-06583-108  
; Sequence 108, Application PC/TUS9606583  
; GENERAL INFORMATION:  
; APPLICANT: DRAYNA, DENNIS T.  
; APPLICANT: FEDER, JOHN N.  
; APPLICANT: GNIKE, ANDREAS  
; APPLICANT: KIMMEL, BRUCE E.  
; APPLICANT: THOMAS, WINSTON J.  
; APPLICANT: WOLFE, ROGER K.  
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
; TITLE OF INVENTION: HEMOCHROMATOSIS  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
; CITY: Washington

STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06583  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,252  
FILING DATE: 09-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 108:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 487 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US96-06583-108

Query Match 1.7% Score 23; DB 5; Length 487;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1227 CACACACACACACACAAAT 1249  
|||||  
Db 322 CACACACACACACACAAAT 344

RESULT 13  
US-08-599-252-102  
Sequence 102, Application US/08599252  
Patent No. 5705343  
GENERAL INFORMATION:  
APPLICANT: DRAVNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GIMKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/599,252  
FILING DATE: 09-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 672 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-599-252-102

Query Match 1.7% Score 23; DB 1; Length 672;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1227 CACACACACACACACAAAT 1249  
|||||  
Db 202 CACACACACACACACAAAT 224

RESULT 14  
PCT-US96-06352-102  
Sequence 102, Application PC/TUS9606352  
GENERAL INFORMATION:  
APPLICANT: DRAVNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GIMKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06352  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,252  
FILING DATE: 09-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 672 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US96-06352-102

Query Match 1.7% Score 23; DB 5; Length 672;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAAT 1249  
|||||  
Db 202 CACACACACACACACAAAT 224

RESULT 15  
PCT-US96-06583-102

; Sequence 102, Application PC/TUS9606583  
; GENERAL INFORMATION:  
; APPLICANT: DRAYNA, DENNIS T.  
; APPLICANT: FEDER, JOHN N.  
; APPLICANT: GRIKE, ANDREAS  
; APPLICANT: KIMMEL, BRUCE E.  
; APPLICANT: THOMAS, WINSTON J.  
; APPLICANT: WOLFE, ROGER K.  
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
; TITLE OF INVENTION: HEMOCHROMATOSIS  
; NUMBER OF SEQUENCES: 124  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/06583  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,252  
; FILING DATE: 09-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 9053-0001.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030  
; INFORMATION FOR SEQ ID NO: 102:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 672 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; PCT-US96-06583-102

Query Match 1.7%; Score 23; DB 5; Length 672;  
Best Local Similarity 100.0%; Pred. No. 0.059;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1227 CACACACACACACACAAAT 1249  
|||||  
Db 202 CACACACACACACACAAAT 224

Search completed: January 18, 2003, 23:37:50  
Job time : 198 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 19:49:34 ; Search time 3606 Seconds

(without alignments)  
11153.654 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382  
Sequence: 1 cattatgctaacaacataa.....taagcaattaagtctctg 1382

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hig:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hig\_hum:\*  
31: em\_hig\_inv:\*  
32: em\_hig\_other:\*  
33: em\_hig\_mus:\*  
34: em\_hig\_pin:\*  
35: em\_hig\_rnd:\*  
36: em\_hig\_mam:\*  
37: em\_hig\_vrt:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1382	100.0	1382	9	AF181722	AF181722 Homo sapi
2	1382	100.0	4377	9	AF181720	AF181720 Homo sapi
3	1382	100.0	120029	2	HSJ82810	AL136722 Homo sapi
4	1352	97.8	152966	9	AL359713	AL359713 Human DNA
5	682	49.3	1287	9	BC014954	BC014954 Homo sapi
6	585	43.1	2167	9	AF181721	AF181721 Homo sapi
7	360	26.0	2003	9	AK027036	AK027036 Homo sapi
8	293	21.2	6548	9	AB032980	AB032980 Homo sapi
9	31	2.2	221641	3	LMPEP1046	AL356683 Leishman
10	29	2.1	38967	3	AC024857	AC024857 Caenorhab
11	29	2.1	41179	3	AF003739	AF003739 Caenorhab
12	29	2.1	49743	2	AC100574	AC100574 Mus muscu
13	29	2.1	110000	2	AC079517_3	Continuation (4 of
14	29	2.1	119318	2	AC097411	AC097411 Rattus no
15	29	2.1	139846	2	AC020895	AC020895 Homo sapi
16	29	2.1	171469	2	AC097344	AC097344 Rattus no
17	29	2.1	201746	2	AC006895	AC006895 Caenorhab
18	29	2.1	223664	2	AC125204	AC125204 Mus muscu
19	29	2.1	236121	2	AC126796	AC126796 Mus muscu
20	28	2.0	497	5	AF254652	AF254652 Tetrao ur
21	28	2.0	47160	2	AC101175	AC101175 Mus muscu
22	28	2.0	48091	2	AC115963	AC115963 Mus muscu
23	28	2.0	78096	2	AC027695	AC027695 Homo sapi
24	28	2.0	82329	9	AL590632	AL590632 Human DNA
25	28	2.0	90111	9	AC008954	AC008954 Homo sapi
26	28	2.0	108175	9	AC068206	AC068206 Homo sapi
27	28	2.0	126355	2	AC126143	AC126143 Rattus no
28	28	2.0	131856	9	AC008971	AC008971 Homo sapi
29	28	2.0	132496	2	AC091841	AC091841 Homo sapi
30	28	2.0	140854	9	AC108038	AC108038 Homo sapi
31	28	2.0	148367	2	AC107744	AC107744 Mus muscu
32	28	2.0	165199	2	AC003973	AC003973 Homo sapi
33	28	2.0	176733	2	AC097686	AC097686 Rattus no
34	28	2.0	180003	2	AC094514	AC094514 Rattus no
35	28	2.0	180675	2	AC126486	AC126486 Rattus no
36	28	2.0	189982	10	AL672274	AL672274 Mouse DNA
37	28	2.0	198037	2	AC126523	AC126523 Rattus no
38	28	2.0	202781	2	AC055817	AC055817 Mus muscu
39	28	2.0	219674	2	AL714010	AL714010 Mus muscu
40	28	2.0	231755	2	AC115480	AC115480 Rattus no
41	27	2.0	259	11	G18504	G18504 cow STS BM4
42	27	2.0	279	4	AF480849	AF480849 Mustela v
43	27	2.0	348	11	G47582	G47582 Z25345_1 Ze
44	27	2.0	412	11	G41134	G41134 Z6977 Zebra
45	27	2.0	60521	2	AC105987	AC105987 Mus muscu

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AF181722	AF181722	Homo sapiens R22AS (R22) mRNA, complete cds.	AF181722	AF181722.1	GI:6684531	Homo sapiens	Homo sapiens	Van Den Eynde, B.J., Gaugler, B., Probst-Keppler, M., Michaux, L., Devuyt, O., Lorge, F., Meynants, P. and Boon, T.	A new antigen recognized by cytolytic T lymphocytes on a human	
AF181722	AF181722	1382 bp mRNA linear	AF181722	AF181722.1	GI:6684531	Homo sapiens	Homo sapiens	Van Den Eynde, B.J., Gaugler, B., Probst-Keppler, M., Michaux, L., Devuyt, O., Lorge, F., Meynants, P. and Boon, T.	A new antigen recognized by cytolytic T lymphocytes on a human	

kidney tumor results from reverse strand transcription  
JOURNAL J. Exp. Med. 190 (12), 1793-1800 (1999)  
MEDLINE 20069887  
PUBMED 10601354  
REFERENCE 2 (bases 1 to 1382)  
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilotte,L.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
Avenue Hippocrate, 74, Brussels 1200, Belgium  
FEATURES  
source  
1. 1382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="p22.1"  
1. 1382  
/gene="RU2"  
738..992  
/gene="RU2"  
/codon\_start=1  
/product="RU2AS"  
/protein\_id="AAF23613.1"  
/db\_xref="GI:6684532"  
/translation="MDDDAARVGEVPAVHKHALHDSLRQVAGPAAAHLPMPPE  
QLASRREAPPLSGRPHRTGAGSPPEPTEKUTPYKKE"  
BASE COUNT 355 a 373 c 344 g 310 t  
ORIGIN

Query Match 100.0%; Score 1382; DB 9; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATATGCTAACAGCATAAACATGCGAGGGGGTGGAGCAGGGTCAAAAAGTGAAGTGTG 60  
Db 1 CATATGCTAACAGCATAAACATGCGAGGGGGTGGAGCAGGGTCAAAAAGTGAAGTGTG 60  
QY 61 TCAATTCCTACTGGAATGAAGTGAATATTTAAACAGTACGGGAAATGCACAGCAA 120  
Db 61 TCAATTCCTACTGGAATGAAGTGAATATTTAAACAGTACGGGAAATGCACAGCAA 120  
QY 121 TTTTCTCCTGCTGTCACATATAGTGTCCACACTTGGAGTGAATTTTAAAGATGTTTA 180  
Db 121 TTTTCTCCTGCTGTCACATATAGTGTCCACACTTGGAGTGAATTTTAAAGATGTTTA 180  
QY 181 TTTTAATTTAAAGATGATGATTTCCAGGAAAAAATAAGAAAAAGAAAAAACTG 240  
Db 181 TTTTAATTTAAAGATGATGATTTCCAGGAAAAAATAAGAAAAAGAAAAAACTG 240  
QY 241 AACGAAAAACGAAAGATATCAGTTGGTCACTAACCTTTCGAAAGATACCTTTTATTT 300  
Db 241 AACGAAAAACGAAAGATATCAGTTGGTCACTAACCTTTCGAAAGATACCTTTTATTT 300  
QY 301 TCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCTGCTGACCCAG 360  
Db 301 TCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCTGCTGACCCAG 360  
QY 361 TGAATTCCTTCTCCAGTCACAGTCAACCTTACCCCCCACTGCAACGAGATTTTG 420  
Db 361 TGAATTCCTTCTCCAGTCACAGTCAACCTTACCCCCCACTGCAACGAGATTTTG 420  
QY 421 AGGGGATCAATCAACAGGAGATGACAGCCCTCAACCACTGAGGTGTGGGGGGTGA 480  
Db 421 AGGGGATCAATCAACAGGAGATGACAGCCCTCAACCACTGAGGTGTGGGGGGTGA 480  
QY 481 GGATCTGATTTCTTCATATCAACCCACACTATAGGGGACCTTAATGGGTGGGGTGG 540  
Db 481 GGATCTGATTTCTTCATATCAACCCACACTATAGGGGACCTTAATGGGTGGGGTGG 540  
QY 541 GGGAGACGACTCACTTGAAGTTCTTTGAAGGCTTCTGAGCTCCAGCCAGTAATTGCC 600  
Db 541 GGGAGACGACTCACTTGAAGTTCTTTGAAGGCTTCTGAGCTCCAGCCAGTAATTGCC 600  
QY 601 CCGCTCTGGATCTGTAGCTTCGGGATTCGGGTGGCCAGTCCGGGGGTGATGATGTTCC 660

Db 601 CCGCTCTGGATCTGTAGCTTCGGGATTCGGGTGGCCAGTCCGGGGGTGATGATGTTCC 660  
QY 661 CTGACGGCCCCAAAGGAGTGCCTGAACGCGCGGCTCACCTCTTCAGAGACTTTCGAAG 720  
Db 661 CTGACGGCCCCAAAGGAGTGCCTGAACGCGCGGCTCACCTCTTCAGAGACTTTCGAAG 720  
QY 721 CTGACACCTTCTTCTCATTCGATGACAGACGGCGCCGCCGCTAGAGAGGCTCCCTTG 780  
Db 721 CTGACACCTTCTTCTCATTCGATGACAGACGGCGCCGCCGCTAGAGAGGCTCCCTTG 780  
QY 781 CGGTACACAGACGCTCTTCACGAGGGGCTAGACAGTGGCTGAGCTGGCTGCTG 840  
Db 781 CGGTACACAGACGCTCTTCACGAGGGGCTAGACAGTGGCTGAGCTGGCTGCTG 840  
QY 841 CCGCTCATCTTCCCGGCTGCGCGCGGCTCAGCTGCTGCTTCCGCTGCGGAGGACCTC 900  
Db 841 CCGCTCATCTTCCCGGCTGCGCGCGGCTCAGCTGCTGCTTCCGCTGCGGAGGACCTC 900  
QY 901 CGCTGCTCCAGCGGCTCTACCGCACCGGAGGCGGGAATCGCTCTCTGAACGAAACGAGA 960  
Db 901 CGCTGCTCCAGCGGCTCTACCGCACCGGAGGCGGGAATCGCTCTCTGAACGAAACGAGA 960  
QY 961 AACTGACGAATCCACAGGTGAAGAGAGTAACGGCGCTGAGCGCTCCACCCAGA 1020  
Db 961 AACTGACGAATCCACAGGTGAAGAGAGTAACGGCGCTGAGCGCTCCACCCAGA 1020  
QY 1021 GGAGACACTAGAGAGCTTCAGAGACTCGGAGTGAAGCTCAAGTTTTCACCGTGGCTGC 1080  
Db 1021 GGAGACACTAGAGAGCTTCAGAGACTCGGAGTGAAGCTCAAGTTTTCACCGTGGCTGC 1080  
QY 1081 ACAGCCAATCAGAGACCCGAGTGGCGCACACACACAGCTTACCTGCTACGGGCAAT 1140  
Db 1081 ACAGCCAATCAGAGACCCGAGTGGCGCACACACAGCTTACCTGCTACGGGCAAT 1140  
QY 1141 CAAGGTGACAGCTCTCTAGAGCAGAGCGGAAAGCGCGGGGCTTCAACAGGACGCGC 1200  
Db 1141 CAAGGTGACAGCTCTCTAGAGCAGAGCGGAAAGCGCGGGGCTTCAACAGGACGCGC 1200  
QY 1201 TAGTGAGGGCAGAGAGAGAGAGAGGAGCAGACACACACACACAAATATGTTGAAACC 1260  
Db 1201 TAGTGAGGGCAGAGAGAGAGAGAGGAGCAGACACACACACAAATATGTTGAAACC 1260  
QY 1261 CAATTTCTTACATCATATCTGCTGCTACCTTTCCAAAGGCTAATTTTCTTTCTCTC 1320  
Db 1261 CAATTTCTTACATCATATCTGCTGCTACCTTTCCAAAGGCTAATTTTCTTTCTCTC 1320  
QY 1321 TTTCTGCACTTTTACCCCTCAATCTCTGCTCTCCCAATTAAGCAATTAAGTTCT 1380  
Db 1321 TTTCTGCACTTTTACCCCTCAATCTCTGCTCTCCCAATTAAGCAATTAAGTTCT 1380  
QY 1381 GG 1382  
Db 1381 GG 1382

RESULT 2  
AF181720/c 4377 bp DNA linear PRI 10-JAN-2000  
LOCUS AF181720 Homo sapiens RU2AS (RU2) gene, complete cds; and RU2S (RU2) gene,  
DEFINITION partial cds.  
ACCESSION AF181720  
VERSION AF181720.1 GI:6684526  
KEYWORDS  
SOURCE Homo sapiens;  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 4377)  
AUTHORS Van den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L.,  
Devuyt,O., Lorge,F., Weynants,P. and Boon,T.  
TITLE A new antigen recognized by cytolytic T lymphocytes on a human  
kidney tumor results from reverse strand transcription

JOURNAL J. Exp. Med. 190 (12), 1793-1800 (1999)  
MEDLINE 20069887  
PUBMED 10601354  
REFERENCE 2 (bases 1 to 4377)  
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilote,L.  
TITLE Direct Submision  
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
Avenue Hippocrate, 74, Brussels 1200, Belgium  
FEATURES Location/Qualifiers  
SOURCE 1. 4377  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p22.1"  
1754..>2348  
/gene="RU2"  
1754..>2348  
/gene="RU2"  
/product="RU2S"  
complement(<1912..>2166)  
/product="RU2AS"  
complement(1912..2166)  
/codon\_start=1  
/product="RU2AS"  
/protein\_id="AAF23611.1"  
/db\_xref="GI:6684528"  
/translation="MDDDAARVEVPYAVHKHLDGLROYAGRAAAHLPRPMP  
QLAASRRAPLPLOSRRHRTGASPPETNEKLTNPQVEK"  
2056..>2348  
/gene="RU2"  
/codon\_start=1  
/product="RU2S"  
/protein\_id="AAF23610.1"  
/db\_xref="GI:6684527"  
/translation="MSGSSARSHLSPVKSVLVYRNGDPFYARRVYIEKKVSE  
EVFLKEVGVGQAFGAVRNITYPTGTGHRIRKLDIOGSGNVVAGGDAFKL"  
BASE COUNT 1296 a 858 c 828 g 1380 t 15 others  
ORIGIN  
Query Match 100.0%; Score 1382; DB 9; Length 4377;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CATATGCTAACAGATTAACATGACAGGGGTGGAGCGAGGTCCAAAGATGAGTTG 60  
DB 2903 CATATGCTAACAGATTAACATGACAGGGGTGGAGCGAGGTCCAAAGATGAGTTG 2844  
QY 61 TCAATTTCTACTGGAATGAAGTTGAATTAATTAACAGTACGGGAATGCAGAGCAA 120  
DB 2843 TCAATTTCTACTGGAATGAAGTTGAATTAATTAACAGTACGGGAATGCAGAGCAA 2784  
QY 121 TTTTCTCTCTGGTGACAAATATAGTGTCCACACTTGAAGTGAATTTTAAAGATGTTA 180  
DB 2783 TTTTCTCTCTGGTGACAAATATAGTGTCCACACTTGAAGTGAATTTTAAAGATGTTA 2724  
QY 181 TTTTAATTAAGATGATTTCCCAAGGAAAAAATAAGAAAAAGAAAAAAGAACG 240  
DB 2723 TTTTAATTAAGATGATTTCCCAAGGAAAAAATAAGAAAAAGAAAAAAGAACG 2664  
QY 241 AACGAAAAACGAAAGATATCAGTTGTGCTACTAACCTTGGCAAGATACCTTTTATTT 300  
DB 2663 AACGAAAAACGAAAGATATCAGTTGTGCTACTAACCTTGGCAAGATACCTTTTATTT 2604  
QY 301 TCTTTAAGATTCCTGTGTGTTTATACACAGATTTTAAAGTTTACTCTACTGTGACCAAG 360  
DB 2603 TCTTTAAGATTCCTGTGTGTTTATACACAGATTTTAAAGTTTACTCTACTGTGACCAAG 2544  
QY 361 TGAATTCCTCTCTGACGTCACAGTGTCAACCTTACCCCAACATGCAACGAGAGTTTG 420  
DB 2543 TGAATTCCTCTCTGACGTCACAGTGTCAACCTTACCCCAACATGCAACGAGAGTTTG 2484  
QY 421 AGGGCATCATCACACCGAGAAGTACAGCCCTCAACCACTGAGGTGGGGGGGTAG 480  
|||||

DB 2483 AGGGCATCATCACACCGAGAAGTACAGCCCTCAACCACTGAGGTGGGGGGGTAG 2424  
QY 481 GGATCTGCAATTTCTTATATCAACCCCACTATAGGCACTTAATGGGTGGCGGTGG 540  
DB 2423 GGATCTGCAATTTCTTATATCAACCCCACTATAGGCACTTAATGGGTGGCGGTGG 2364  
QY 541 GGGAGACGACATCTGAGTTTCTTGAAGGGTCTCTGGCCCTCCAGCAAGTAATGGCC 600  
DB 2363 GGGAGACGACATCTGAGTTTCTTGAAGGGTCTCTGGCCCTCCAGCAAGTAATGGCC 2304  
QY 601 CCGCTCTGATCTGTGATAGTACGAGCGGGCGCGGTAGAAAGGGGTCCCGTTG 660  
DB 2303 CCGCTCTGATCTGTGATAGTACGAGCGGGCGCGGTAGAAAGGGGTCCCGTTG 2244  
QY 661 CTGACGGCCCCAAAGGGTCCCTGAAACGCCCGGGTCTACTCTTCAAGAAAGCTTGAAG 720  
DB 2243 CTGACGGCCCCAAAGGGTCCCTGAAACGCCCGGGTCTACTCTTCAAGAAAGCTTGAAG 2184  
QY 721 CTGACACCTTCTTCATGATGAGACGAGCGGGCGCGGTAGAAAGGGGTCCCGTTG 780  
DB 2183 CTGACACCTTCTTCATGATGAGACGAGCGGGCGCGGTAGAAAGGGGTCCCGTTG 2124  
QY 781 CCGTACACAAAGCAGCTCTTCAAGAGCGGCTGAGACAGGTGCTGACCTGGCGCTGCTG 840  
DB 2123 CCGTACACAAAGCAGCTCTTCAAGAGCGGCTGAGACAGGTGCTGACCTGGCGCTGCTG 2064  
QY 841 CCGCTCATCTTCCCGCTGGCGCGCGCTCAGCTGCTGCTGCTGCGTGGAGGCACTTC 900  
DB 2063 CCGCTCATCTTCCCGCTGGCGCGCGCTCAGCTGCTGCTGCTGCGTGGAGGCACTTC 2004  
QY 901 CCGTCTCCAGCGGCGCTCAGCGACCGAGGGCGGGGATCGCTCTGAAAGCAAGAGA 960  
DB 2003 CCGTCTCCAGCGGCGCTCAGCGACCGAGGGCGGGGATCGCTCTGAAAGCAAGAGA 1944  
QY 961 AACTGACGATCCACAGTGAAGAGAAAGTAACGGCGCTGCGCTAGCGGCTCCACCAGA 1020  
DB 1943 AACTGACGATCCACAGTGAAGAGAAAGTAACGGCGCTGCGCTAGCGGCTCCACCAGA 1884  
QY 1021 GGAAGACCTAAGAGCTTGCAGAGACTGGAGTACGCTTAAGTTTTCACCGTGGCGTGC 1080  
DB 1883 GGAAGACCTAAGAGAGCTTGCAGAGACTGGAGTACGCTTAAGTTTTCACCGTGGCGTGC 1824  
QY 1081 ACAGCCAAATCAGAGACCGGAGTGGCGGCAACACAGGTTCACGTTACCGGGCAAT 1140  
DB 1823 ACAGCCAAATCAGAGACCGGAGTGGCGGCAACACAGGTTCACGTTACCGGGCAAT 1764  
QY 1141 CAAGGTGACAGCTTCTGAGCAGAGACCGGAAACGGCGGGGCTTCAAAACAGGCAAGCC 1200  
DB 1763 CAAGGTGACAGCTTCTGAGCAGAGACCGGAAACGGCGGGGCTTCAAAACAGGCAAGCC 1704  
QY 1201 TAGTGAAGGACAGAGAGAGAGAGCGACACACACACACACAAATATGTGTAACCC 1260  
DB 1703 TAGTGAAGGACAGAGAGAGAGAGAGCGACACACACACACAAATATGTGTAACCC 1644  
QY 1261 CAATTTCTTAACATATCTGTGCTACCGCTTCCAAAGCGCTAATTTTCTTCTCTC 1320  
DB 1643 CAATTTCTTAACATATCTGTGCTACCGCTTCCAAAGCGCTAATTTTCTTCTCTC 1584  
QY 1321 TTTCTGACACTTACCCCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1380  
DB 1583 TTTCTGACACTTACCCCTCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1524  
QY 1381 GG 1382  
DB 1523 GG 1522  
RESULT 3  
HSJ282H10 120029 bp DNA linear HTG 10-JUL-2001  
LOCUS HSJ282H10  
DEFINITION Homo sapiens chromosome 6 clone RP1-282H10 map p22.1-22.3, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 5 unordered pieces.  
ACCESSION AL132672

```

VERSION      AL132672.14  GI:14448905
KEYWORDS     HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 120029)
AUTHORS      Garner, P.
TITLE        Direct Submission
              Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
              CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
              requests: clonerequest@sanger.ac.uk
              On Jun 12, 2001 this sequence version replaced g1:12331282.
COMMENT      ----- Sanger Centre
              Centre genome
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              ----- Project Information
              Center project name: dj282H10
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Sequencing vector: M13; M77815; 0% of reads
              Sequencing vector: plasmid; L08752; 99% of reads
              Chemistry: dye-terminator ABI; 1% of reads
              Chemistry: dye-terminator Big Dye; 98% of reads
              Consensus quality: 118224 bases at least Q40
              Consensus quality: 118449 bases at least Q30
              Consensus quality: 118634 bases at least Q20
              Insert size: 119629; sum-of-contigs
              Insert size: 117431; 9.3% error; agarose-gel
              Quality coverage: 18.41x in Q20 bases; sum-of-contigs
              Quality coverage: 19.36x in Q20 bases; agarose-gel
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 5 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 23429: contig of 23429 bp in length
              * 23430 23528: gap of 100 bp
              * 23530 63600: contig of 40071 bp in length
              * 63601 63700: gap of 100 bp
              * 63701 100099: contig of 36399 bp in length
              * 100100 100199: gap of 100 bp
              * 100200 102366: contig of 2167 bp in length
              * 102367 102466: gap of 100 bp
              * 102467 120029: contig of 17563 bp in length.
              Location/Qualifiers
                1..120029
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="6"
                  /map="P22.1-22.3"
                  /clone="RP1-282H10"
                  /clone.lib="RPC1-1"
                1..23429
                  /note="assembly_fragment:05837
                  clone_end:SP6
                  vector_side:left"
                23530..63600
                  /note="assembly_fragment:00423
                  fragment_chain:1"
                63701..100099
                  /note="assembly_fragment:02685
                  fragment_chain:1"
                100200..102366
                  /note="assembly_fragment:01789"
                102467..120029
                  /note="assembly_fragment:02341

```

BASE COUNT	36483 a	24840 c	24637 g	33666 t	403 others
Query Match	100.0%	Score 1382;	DB 2;	Length 120029;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 1382;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
1	CATTATGCTAACAGATATAACATGACGAGGGGGTGGAGACAGGGGTCTCAAAAGTACGAGTGTG	60			
21048	CATTATGCTAACAGATATAACATGACGAGGGGGTGGAGACAGGGGTCTCAAAAGTACGAGTGTG	21107			
61	TCATTTCTACTTGGAAATGAAGGTTGAATTAATTTAACAGTACGAGGAAATCAGACGAA	120			
21108	TCATTTCTACTTGGAAATGAAGGTTGAATTAATTTAACAGTACGAGGAAATCAGACGAA	21167			
121	TTTTTCCCTCGGTGACATATAGTGTCCACACATTGGAAAGGATTTTAAAGATGTTTA	180			
21168	TTTTTCCCTCGGTGACATATAGTGTCCACACATTGGAAAGGATTTTAAAGATGTTTA	21227			
181	TTTTAATTTAAAGATGATTTTCCAAAGAAAAAAATAAGAAAAAGAAAAAAACCTG	240			
21228	TTTTAATTTAAAGATGATTTTCCAAAGAAAAAAATAAGAAAAAGAAAAAAACCTG	21287			
241	AACAGAAAAACGAAAGTATCATGTTGGTCTAACCTTTGCAAGAGTACCTTTTATTT	300			
21288	AACAGAAAAACGAAAGTATCATGTTGGTCTAACCTTTGCAAGAGTACCTTTTATTT	21347			
301	TCCTTTAAGTTCCTGTGTGTTTATACACAGATTTTAAGTTTACTCCACGTCAGCCCAAG	360			
21348	TCCTTTAAGTTCCTGTGTGTTTATACACAGATTTTAAGTTTACTCCACGTCAGCCCAAG	21407			
361	TGAAATTCCTTCAGTACACAGTGTCAACCTCTACCCGCCCACTGCAACGAGAGTTTGG	420			
21408	TGAAATTCCTTCAGTACACAGTGTCAACCTCTACCCGCCCACTGCAACGAGAGTTTGG	21467			
421	AGGGGATTCATACACCGAGAAATACAGCCCTTCACACTGAGGTGTGGGGGGGTAG	480			
21468	AGGGGATTCATACACCGAGAAATACAGCCCTTCACACTGAGGTGTGGGGGGGTAG	21527			
481	GGATTCGATTCCTTCATTCATTCACACCCCAACATTAAGGAGACCTAAATGGGTGGGGGTG	540			
21528	GGATTCGATTCCTTCATTCATTCACACCCCAACATTAAGGAGACCTAAATGGGTGGGGGTG	21587			
541	GGGAGACGACACTCAGTTGATGATTTCTTGAAGGCTTCCTGCGCTCCAGCCACGTAATTTGCC	600			
21588	GGGAGACGACACTCAGTTGATGATTTCTTGAAGGCTTCCTGCGCTCCAGCCACGTAATTTGCC	21647			
601	CCGCTTCGATTCGTGCTTACGATTCGGGATTCGGTGGCCAGTCCCGGGGGTGTAGATTTTC	660			
21648	CCGCTTCGATTCGTGCTTACGATTCGGGATTCGGTGGCCAGTCCCGGGGGTGTAGATTTTC	21707			
661	CTGAGGGCCCCCAAGAGGTGCTGAAACCGCGCGGTACCTCTTCAGAGAACTTCGAAG	720			
21708	CTGAGGGCCCCCAAGAGGTGCTGAAACCGCGCGGTACCTCTTCAGAGAACTTCGAAG	21767			
721	CTGAGACACCTTCTTCATGATGATGACAGCGCGCGCCCGCTAGAGAGGGGTCCCGGTGG	780			
21768	CTGAGACACCTTCTTCATGATGATGACAGCGCGCGCCCGCTAGAGAGGGGTCCCGGTGG	21827			
781	CGGTACACAGACGCTTTCACAGACGGGCTGAGACAGGTGGCTGGACCTGGCGGTGCTG	840			
21828	CGGTACACAGACGCTTTCACAGACGGGCTGAGACAGGTGGCTGGACCTGGCGGTGCTG	21887			
841	CCGCTCATCTTCCCGGTGGCGCGCGCTCAGCTCGCTGCTTCGCGTGGGAGGACACTTC	900			
21888	CCGCTCATCTTCCCGGTGGCGCGCGCTCAGCTCGCTGCTTCGCGTGGGAGGACACTTC	21947			
901	CGCTGTCCACAGGGGCTCACCGCACCCACAGGGGGGGGATCGGCTCTGAAAGCAAGACGAGA	960			
21948	CGCTGTCCACAGGGGCTCACCGCACCCACAGGGGGGGGATCGGCTCTGAAAGCAAGACGAGA	22007			

QY	961	AACTACGCAATTCACAGGTGAAAGAAAGTAACGGCCCTGGCCTTAGGCCTCACCCAGA	1020
Db	22008	AACTGACCAATTCACAGGTGAAAGAAAGTAACGGCCCTGGCCTTAGGCCTCACCCAGA	22067
QY	1021	GGAGACACTAGGAGCTTCAGAGACTCGGAGATGAGCGTCACAGTTTTTTCACCGCTGGCGAGC	1080
Db	22068	GGAGACACTAGGAGCTTCAGAGACTCGGAGATGAGCGCTCMAAGTTTTTTCACCGCTGGCGAGC	22127
QY	1081	ACAGCCAATTCAGGACCGCGCAGTGGCGGACACACACAGGTTTCACCTGGTACGGGAGAA	1140
Db	22128	ACAGCCAATTCAGGACCGCGCAGTGGCGGACACACACAGGTTTCACCTGGTACGGGAGAA	22187
QY	1141	CAAGGTGACAGCTTCTAGAGAGAGGCGGGGAAAGCGGGGGGCTTCAAACAGGACGCC	1200
Db	22188	CAAGGTGACAGCTTCTAGAGAGAGGCGGGGAAAGCGGGGGGCTTCAAACAGGACGCC	22247
QY	1201	TAGTGAGGCGCAGAGAGAGAGGAGGACACACACACACACAAATATGSGTAAACC	1260
Db	22248	TAGTGAGGCGCAGAGAGAGAGGAGGACACACACACACACAAATATGSGTAAACC	22307
QY	1261	CAATTTCTTAAATCTATATCTGTGCTAACCCCTTCCAAAGAGCTAATTTTTCTTTCTCTC	1320
Db	22308	CAATTTCTTAAATCTATATCTGTGCTAACCCCTTCCAAAGAGCTAATTTTTCTTTCTCTC	22367
QY	1321	TTCTTGACACTTTAACCCCTCAATCTCGTGTCTCCGCCAAATTTAAAGCAATTAAGTTTCC	1380
Db	22368	TTCTTGACACTTTAACCCCTCAATCTCGTGTCTCCGCCAAATTTAAAGCAATTAAGTTTCC	22427
QY	1381	GG 1382	
Db	22428	GG 22429	

RESULT 4	AL359713	152966 bp	DNA	linear	PRI 01-MAY-2001
LOCUS	AL359713				
DEFINITION	Human DNA sequence from clone RP11-95P3 on chromosome 6, complete sequence.				
ACCESSION	AL359713				
VERSION	AL359713.25				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 152966)				
TITLE	Kimberley/A.				
JOURNAL	Direct Submission				
COMMENT	Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Requests: clonequests@sanger.ac.uk On May 3, 2001 this sequence version replaced gi:13446455. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at				

<http://www.sanger.ac.uk/HGP/Chr6>  
 RP11-95P3 is from the library RP01-11.1 constructed by the group of  
 Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBAC3.6  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP11-95P3. It may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone RP11-95P3 is at 1 in this sequence. The  
 true left end of clone RP1-73M23 is at 152867 in this sequence. The  
 true right end of clone RP11-40E20 is at 17700 in this sequence.  
 Location/Qualifiers  
     1..152966

misc_feature	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /clone_idb="RPC1-11.1" 19..544 /note="match: GSS: Em:A2517849" 24..379
misc_feature	/note="match: GSS: Em:B65526" 32..494 /note="match: GSS: Em:AQ315706" 70..589 /note="match: GSS: Em:AQ285677" 179..362
repeat_region	/note="MIR repeat: matches 65..262 of consensus" 521..582 /note="L2 repeat: matches 2683..2741 of consensus" 720..1022
repeat_region	/note="Alusx repeat: matches 1..303 of consensus" 1327..1448 /note="AluG/PRAM repeat: matches 181..298 of consensus" 2264..2575 /note="AlusG repeat: matches 1..311 of consensus" 3014..3296 /note="match: STS: Em:G27290" 3014..3093 /note="40 copies 2 mer ga 75% conserved" 5024..5477
repeat_region	/note="L1M1 repeat: matches 1012..1598 of consensus" 5581..5859 /note="AlusG repeat: matches 18..294 of consensus" 5986..6359 /note="L1Mc repeat: matches 2228..2259 of consensus" 6367..6515 /note="FLAM_C repeat: matches 1..143 of consensus" 6971..7198 /note="L2 repeat: matches 1512..1735 of consensus" 8403..8838 /note="LRR7 repeat: matches 1..445 of consensus" 8799..9441 /note="match: GSS: Em:B54691" complement(9393..9853) /note="match: GSS: Em:AQ285817" 10311..10437
repeat_region	/note="L2 repeat: matches 2612..2747 of consensus" 10710..11022 /note="AluG repeat: matches 1..309 of consensus" 12006..12156 /note="L2 repeat: matches 2594..2748 of consensus" 12182..13011 /note="L2 repeat: matches 1374..2344 of consensus" 13404..13631 /note="L14 copies 2 mer at 55% conserved" 13479..13630 /note="38 copies 4 mer tata 60% conserved" 13785..14820 /note="L1Mc repeat: matches 4766..5798 of consensus" 14928..16008 /note="L1P3 repeat: matches 5106..6149 of consensus" 16029..16110
repeat_region	

	repeat_region	/note="AluX repeat: matches 1. .297 of consensus" 34621. .34732	
	repeat_region	/note="MIR repeat: matches 120. .232 of consensus" 34888. .35007	
	repeat_region	/note="L2 repeat: matches 1836. .1959 of consensus" 35164. .35461	
	repeat_region	/note="AluG repeat: matches 2. .298 of consensus" 35352. .35990	
	repeat_region	/note="L2 repeat: matches 2254. .2710 of consensus" 36631. .36790	
	repeat_region	/note="L2 repeat: matches 2242. .2501 of consensus" 37251. .37544	
	repeat_region	/note="AluX repeat: matches 1. .294 of consensus" 38446. .38493	
	repeat_region	/note="L2 copies 4 mer caca 75% conserved" 38467. .38492	
	Query Match	97.8%; Score 1352; DB 9; Length 152966;	
	Best Local Similarity	100.0%; Pore No. 0;	
	Matches 1352; Conservative	0; Mismatches 0; Indels 0; Gaps	
QY	1	CATTTCGTAAACAGATATAACATGACGAGGGGAGGAGCAGGGCTCAACAAAGTAGAGTTG	60
DB	53422	CATTTCGTAAACAGATATAACATGACGAGGGGAGGAGCAGGGCTCAACAAAGTAGAGTTG	53488
QY	61	TCAAATTCACCTTGAATGAAAGGTTGAAATTAATTAACAGTACGGGAAATCAGAGCAA	120
DB	53482	TCAAATTCACCTTGAATGAAAGGTTGAAATTAATTAACAGTACGGGAAATCAGAGCAA	53541
QY	121	TTTTTCCTCCCTGGTACATATAGTGTCCACACCTTGGAAGGATTTTAAACAATCTTAA	180
DB	53542	TTTTTCCTCCCTGGTACATATAGTGTCCACACCTTGGAAGGATTTTAAACAATCTTAA	53601
QY	181	TTTAAATTTAAAGATGATGATTTCCAAAGAAAAAATAAGAAAAAGAAAAAACTG	240
DB	53602	TTTAAATTTAAAGATGATGATTTCCAAAGAAAAAATAAGAAAAAGAAAAAACTG	53665
QY	241	AACAGAAAAAGCAAAAGTACAGTTTGGTCTAACCTTTGCAAGGATACCTTTTATTT	300
DB	53662	AACAGAAAAAGCAAAAGTACAGTTTGGTCTAACCTTTGCAAGGATACCTTTTATTT	53722
QY	301	TCTTTAAGATTCCTGTGTGTTTATACACAGATTTTAAGTTTACTCCACGCTGACCCAG	360
DB	53722	TCTTTAAGATTCCTGTGTGTTTATACACAGATTTTAAGTTTACTCCACGCTGACCCAG	53781
QY	361	TGAATTCCTCTCCAGTACAGTCACTCTACCCCCCAACTGCAAGAGAGTTTG	420
DB	53782	TGAATTCCTCTCCAGTACAGTCACTCTACCCCCCAACTGCAAGAGAGTTTG	53844
QY	421	AGGGGCATCAATCACACCGAAGTACAGCCCTCAACCACTGAGGTGGGGGGGTAG	480
DB	53842	AGGGGCATCAATCACACCGAAGTACAGCCCTCAACCACTGAGGTGGGGGGGTAG	53901
QY	481	GGATTCGATCTTCTTCATATCAACCCACACATATAGGGCACTAAATGGGTGGGGGTG	540
DB	53902	GGATTCGATCTTCTTCATATCAACCCACACATATAGGGCACTAAATGGGTGGGGGTG	53961
QY	541	GGGAGACGACACACTGTGTTCTTGAAGGGTCTCTGGCCCTCAAGCCAGTAATTGGCC	600
DB	53962	GGGAGACGACACACTGTGTTCTTGAAGGGTCTCTGGCCCTCAAGCCAGTAATTGGCC	54021
QY	601	CCGCTCTGGATCTGGTCTAGTTCGAGTCCGGATTCGGTGGCCAGTCCGGGGGTGATGATGTC	660
DB	54022	CCGCTCTGGATCTGGTCTAGTTCGAGTTCGGATTCGGTGGCCAGTCCGGGGGTGATGATGTC	54081
QY	661	CTGAGCGCCCCAAAGGGTGCCTGAAGCGCCGCTCACCTCTTCAGAAAGACTTCGAAG	720
DB	54082	CTGAGCGCCCCAAAGGGTGCCTGAAGCGCCGCTCACCTCTTCAGAAAGACTTCGAAG	54141
QY	721	CTGAGCACCTTTCTCATGATGACAGCGGGGCCCCGGCTAAGAGGGGTCCCGGTG	780
DB	54142	CTGAGCACCTTTCTCATGATGACAGCGGGGCCCCGGCTAAGAGGGGTCCCGGTG	54201

```

QY 781 CGGTACACAGACGCTTTCACGAGCGGCTGAGAGGTGCTGACCTGCGCTGCTG 840
    |||||||
Db 54202 CGGTACACAGACGCTTTCACGAGCGGCTGAGAGGTGCTGACCTGCGGCTGCTG 54261
    |||||||
QY 841 CCGCTCATCTTCCCGCGCTGCGCGCGCTGAGCTGCTGCTGCTGCTGCGGAGGACCTTC 900
    |||||||
Db 54262 CCGCTCATCTTCCCGCGCTGCGCGCGCTGAGCTGCTGCTGCTGCTGCGGAGGACCTTC 54321
    |||||||
QY 901 CCGCTGCCAGGCGGCTGACCGGACCGGAGGCGCGGAGATGCTCTCTGAAACGACGAGA 960
    |||||||
Db 54322 CCGCTGCCAGGCGGCTGACCGGACCGGAGGCGCGGAGATGCTCTCTGAAACGACGAGA 54381
    |||||||
QY 961 AACTGACGAATCCACAGAGGAAAGAGAATTAACGGCGCTGCGCTGAGCGCTCCACCAGA 1020
    |||||||
Db 54382 AACTGACGAATCCACAGAGGAAAGAGAATTAACGGCGCTGCGCTGAGCGCTCCACCAGA 54441
    |||||||
QY 1021 GGAGACACTAGAGAGCTTGGAGAGAGCTGGAGTAGAGCTCAAGTTTTCACCGGTGGCGTGC 1080
    |||||||
Db 54442 GGAGACACTAGAGAGCTTGGAGAGAGCTGGAGTAGAGCTCAAGTTTTCACCGGTGGCGTGC 54501
    |||||||
QY 1081 AAGACCAATCAGAGACCGCGAGTCCGCGACACACAGGTTCACCTGCTACGGGCGAGAAT 1140
    |||||||
Db 54502 AAGACCAATCAGAGACCGCGAGTCCGCGACACACAGGTTCACCTGCTACGGGCGAGAAT 54561
    |||||||
QY 1141 CAAGTGAGACAGCTTCTGAGACAGAGCCGGAACGCGCGGCGCTCAACAGAGGACGCG 1200
    |||||||
Db 54562 CAAGTGAGACAGCTTCTGAGACAGAGCCGGAACGCGCGGCGCTCAACAGAGGACGCG 54621
    |||||||
QY 1201 TTAGTGAGGCGAGAGAGAGAGAGCGACACACACACACACAAATATGTTGTAAC 1260
    |||||||
Db 54622 TTAGTGAGGCGAGAGAGAGAGAGCGACACACACACACACAAATATGTTGTAAC 54681
    |||||||
QY 1261 CAATTTCTACATCATATGCTGCTACCTTTCCAAACAGCCCTAATTTTCTTTCTCTC 1320
    |||||||
Db 54682 CAATTTCTACATCATATGCTGCTACCTTTCCAAACAGCCCTAATTTTCTTTCTCTC 54741
    |||||||
QY 1321 TTCTTGACCTTTTACCCCTCAATCTCTGCTT 1352
    |||||||
Db 54742 TTCTTGACCTTTTACCCCTCAATCTCTGCTT 54773
    |||||||

RESULT 5
BC014954/c 1287 bp mRNA linear PRI 04-OCT-2001
LOCUS BC014954
DEFINITION Homo sapiens, clone MGC:22980 IMAGE:4874845, mRNA, complete cds.
ACCESSION BC014954
VERSION BC014954.1 GI:15928979
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1287)
Strausberg, R.
Direct Submission
Submitted (01-OCT-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NHI-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice Mcleavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhnu, Parvaneh Saeedi, Jacqueline

```

```

FEATURES
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="MGC:22980 IMAGE:4874845"
        /tissue_type="Eye, normal, pigmented retinal epithelium"
        /clone_1lb="NHI-MGC_43"
        /lab_host="DH10B-R"
        /note="Vector: pOTB7"
        181..984
    CDS
        /codon_start=1
        /product="Unknown (protein for MGC:22980)"
        /protein_id="AA14954.1"
        /db_xref="GI:15928980"
        /translation="MGAVASAGEDNDLDLNLKEAOYIFTEREEOAFRAIDRGDYILE
        GYRDNAVXKDLAKHGNHILSAPCIYSEVWEALKLPGLSFLNIGSGTGLSTMVGLIL
        GPFGIHGIEHSDYVEAKERLESFDRSRSLGAVRPIGRCRLPTRSS
        ELRRRPAKMSSSSARSSHLSPYKSVLYVRNGPFRVYIHEKRVSFYVFLK
        EVTGVPQAFGAVRNVIYTPTRHRLKLDQIOSGGVYAGVGEAFKRLK"
BASE COUNT      343 a      281 c      359 g      304 t
ORIGIN
Query Match      49.3% Score 682; DB 9; Length 1287;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;
Matches 682; Conservative 0;
QY 264 TTGTGTCACATTAACCTTTGCAAGATACCTTTTATTTTCTTAAAGATTCCTGTTGTTAT 323
    |||||||
Db 1272 TTGTGTCACATTAACCTTTGCAAGATACCTTTTATTTTCTTAAAGATTCCTGTTGTTAT 1213
    |||||||
QY 324 ACACAGATTTTAAAGTTTACTTCTCTACCTGCTGACCCAAAGTAAATTCCTTCCAGTACAG 383
    |||||||
Db 1212 ACACAGATTTTAAAGTTTACTTCTCTACCTGCTGACCCAAAGTAAATTCCTTCCAGTACAG 1153
    |||||||
QY 384 TGCACACCTCTACCCGCCCAACGACGAGAGTTTGGAGGGCATCAATACACCGAGAA 443
    |||||||
Db 1152 TGCACACCTCTACCCGCCCAACGACGAGAGTTTGGAGGGCATCAATACACCGAGAA 1093
    |||||||
QY 444 GTCCACAGCCCTTCACACCTAGAGTGTGGGGGGTGAAGATTCGATTTCTTCATATCAA 503
    |||||||
Db 1092 GTCCACAGCCCTTCACACCTAGAGTGTGGGGGGTGAAGATTCGATTTCTTCATATCAA 1033
    |||||||
QY 504 CCCACACATATAGGACACCTAAATGGGTGGCGGTGGGGAGACCGACCTGACCTTGA 563
    |||||||
Db 1032 CCCACACATATAGGACACCTAAATGGGTGGCGGTGGGGAGACCGACCTGACCTTGA 973
    |||||||
QY 564 CTGGAAGGCTCTCTGGGCTTCACGACGATATGGCCCCCGCTGTGGATGCTGCTGAGCTT 623
    |||||||
Db 972 CTGGAAGGCTCTCTGGGCTTCACGACGATATGGCCCCCGCTGTGGATGCTGCTGAGCTT 913
    |||||||
QY 624 CCGGATTCGGTGGCCAGTCCGCGGGGTGTAGATGTTCTGACGCGCCCAAGGGTGCCTG 683
    |||||||
Db 912 CCGGATTCGGTGGCCAGTCCGCGGGGTGTAGATGTTCTGACGCGCCCAAGGGTGCCTG 853
    |||||||
QY 684 AACGCGCGGGGTGACCTCTTCAGGAAGACTTGAAGCTGGACACCTTCTTCATGAGAT 743
    |||||||
Db 852 AACGCGCGGGGTGACCTCTTCAGGAAGACTTGAAGCTGGACACCTTCTTCATGAGAT 793
    |||||||
QY 744 GACGACGCGGCGCCGCGGATGAGAAAGGGTCCCGTTGGCGGTACACAAGACGCTTTCAC 803
    |||||||
Db 792 GACGACGCGGCGCCGCGGATGAGAAAGGGTCCCGTTGGCGGTACACAAGACGCTTTCAC 753
    |||||||
QY 804 GACGCGCTGAGACAGGTGCTGACCTGGCGCTGCTGCGCTCATCTTCCCGCTGCGCG 863
    |||||||

```

```

Db 732 GACGGGCTGAGACAGGTGGTGGACCTGGCGTGGCTGCATCTTCCCGCTGAGCG 673
QY 864 CCGGCTCAGCTCGCTGCTTCCGTCGGGAGGACACCTCCGCTGTCCACGCGGCTCAGCCG 923
Db 672 CCGGCTCAGCTCGCTGCTTCCGTCGGGAGGACACCTCCGCTGTCCACGCGGCTCAGCCG 613
QY 924 ACCCAGGGGCGGGGATCGCTC 945
Db 612 ACCCAGGGGCGGGGATCGCTC 591

RESULT 6
LOCUS AF181721 2167 bp mRNA linear PRI 10-JAN-2000
DEFINITION Homo sapiens Ruz2 (RUZ) mRNA, complete cds.
ACCESSION AF181721
VERSION AF181721.1 GI:6684529
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 2167)
AUTHORS Van den Eynde,B.,J., Gaugler,B., Probst-Kepper,M., Michaux,L.,
Devuyst,O., Lorge,F., Weynants,P. and Boon,T.
TITLE A new antigen recognized by cytolytic T lymphocytes on a human
JOURNAL kidney tumor results from reverse strand transcription
MEDLINE J Exp. Med. 190 (12), 1793-1800 (1999)
PUBMED 10601354
REFERENCE 2 (bases 1 to 2167)
AUTHORS Van den Eynde,B.,J., Gaugler,B. and Pilotte,L.
TITLE Direct Submission
JOURNAL Submitted (30-ANG-1999) Ludwig Institute for Cancer Research,
FEATURES source
source Location/Qualifiers
1..2167
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p22.1"
1..2167
/gene="RU2"
303..1733
/gene="RU2"
/codon_start=1
/product="RU2S"
/protein_id="AA23612.1"
/db_xref="GI:6684530"
/translacion="MSGSSARSSHLQPVVKSVLVYRNGDPFYAGRRVITHEKVSF
EVLKEVTGVOAPGAVRNITPRTGHRIRKLDIOISGNYVAGGEVAFKRLKLYLDI
GEIKRREVVNTEVKVIRSRINVASRFRKPEPTITLIANGDILNARSLLIIR
KITLONPHVLOMTTEKITLRSQAVHRLYTEGLKVESGAELENGQFYVAVGRDFKL
PYGELLEFDKSTMRPFQKASLPPYVSGSKSGSGMDRSTRKSTVSGSDNSDPPLK
KGRKEDVNSKEITLKNQVKNKINSQETIPNSDEGIFKAGAESETRGAARVQEDPD
VEVPVDQPAEIVDEDEDEKANKADEOKEDFSGMDLEEGEGREATDAPQVEEIL
DHSEQOARPARVNGSGTDEENGEELOVNNELQVLDKERSQSGASQDEADVDPPQR
PREPVKTSPEENNNQNKDYAVNA"
BASE COUNT 675 a 440 c 576 g 476 t
ORIGIN
Query Match 43.1%; Score 595; DB 9; Length 2167;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 676 GGTGCTGAACCGCGCCGCTACCTCCTTACAGAAAGACTTGAAGCTGAGACCTTCTTC 735
Db 475 GGTGCTGAACCGCGCCGCTACCTCCTTACAGAAAGACTTGAAGCTGAGACCTTCTTC 416
QY 736 TCATGATGACGACGCGCGCGCGCGCTGAGAAAGGGGTCCCGTTGCGGTACACAAGACG 795
Db 415 TCATGATGACGACGCGCGCGCGCGCTGAGAAAGGGGTCCCGTTGCGGTACACAAGACG 356
QY 796 CTCTTCACGACGCGCGCTGAGACAGGTGGTGGACCTGGGCGTGGCGCTCATCTTCCG 855
Db 355 CTCTTCACGACGCGCGCTGAGACAGGTGGTGGACCTGGGCGTGGCGCTCATCTTCCG 296
QY 856 GCTGCGCGCGCTCAGCTCGCTGCTTCCGTCGGGAGGACCTCCGCTGCCAGCGGC 915
Db 295 GCTGCGCGCGCTCAGCTCGCTGCTTCCGTCGGGAGGACCTCCGCTGCCAGCGGC 236
QY 916 CTCACCGCACCAGCGCGCGCGGATCGCTCTGAAACGACGAGAACTACGATCCAC 975
Db 235 CTCACCGCACCAGCGCGCGCGGATCGCTCTGAAACGACGAGAACTACGATCCAC 176
QY 976 AGGTGAAGAGAAAGTAAACGGCGCTGAGGGCTCCACCCAGAGAGACATAGAGAC 1035
Db 175 AGGTGAAGAGAAAGTAAACGGCGCTGAGGGCTCCACCCAGAGAGACATAGAGAGC 116
QY 1036 TTGCAGAGCTGGAGTAGACGCTCAAGTTTTCACCGGTGGGTCACAGCAATCAGAC 1095
Db 115 TTGCAGAGCTGGAGTAGACGCTCAAGTTTTCACCGGTGGGTCACAGCAATCAGAGC 56
QY 1096 CCGCAGTCCGCGCACACACACAGGTTCCACGTCTACGGGCGAATCAAGGTGAC 1150
Db 55 CCGCAGTCCGCGCACACACACAGGTTCCACGTCTACGGGCGAATCAAGGTGAC 1

RESULT 7
LOCUS AK027036 2003 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23383 fis, clone HEP16466, highly similar to
AF181721 Homo sapiens Ruz2 mRNA.
ACCESSION AK027036
VERSION AK027036.1 GI:10440050
KEYWORDS Oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens hepatoma cell_line:Hep2 cDNA to mRNA, clone_11b:HEP
clone:HEP16466.
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Odayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
2 (bases 1 to 2003)
Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shitokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA library
sequencing: Research Association for Biotechnology; cDNA library
construction, 5' - 8' 3'-end one pass sequencing: Departent of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES source
source Location/Qualifiers
1..2003
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HEP16466"

```



```

/misc.feature      1..2003
                   /note="highly similar to AF181721 Homo sapiens RUS2 mRNA"
BASE COUNT        662 a      388 c      525 g      428 t
ORIGIN

Query Match      26.0%  Score 360;  DB 9;  Length 2003;
Best Local Similarity 99.8%  Pred. No. 1.5e-204;
Matches 410;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

556 TTGAGTTCTTGAAGGCTTCTGGCCCTCCAGCAGCAAGTAATTCGCCCGCTCGATCTGG 615
    |||||||
Db 411 TTGAGTTCTTGAAGGCTTCTGGCCCTCCAGCAGCAAGTAATTCGCCCGCTCGATCTGG 352

QY 616 TCTAGCTTCGGATTCCGTGGCCAGTCCGCGGGGTGTAGATTCTCTGACGCGCCCAAG 675
    |||||||
Db 351 TCTAGCTTCGGATTCCGTGGCCAGTCCGCGGGGTGTAGATTCTCTGACGCGCCCAAG 292

QY 676 GGTGCTTGAAGCGCGCGGTCTACCTCTTCAAGAAAGATTGGAAGCTGACACCTTCTTC 735
    |||||||
Db 291 GGTGCTTGAAGCGCGCGGTCTACCTCTTCAAGAAAGATTGGAAGCTGACACCTTCTTC 232

QY 736 TCATGATGACGACGCGCGCGCGCGGTGAGAAAGGCGTCCCGTTGCGGTACAAAGACAG 795
    |||||||
Db 231 TCATGATGACGACGCGCGCGCGCGGTGAGAAAGGCGTCCCGTTGCGGTACAAAGACAG 172

QY 796 CTCCTTACGACGACGCGCGCGGTGAGAGTGGCTGACCTGCGGCTGCTCATCTTCC 855
    |||||||
Db 171 CTCCTTACGACGACGCGCGGTGAGAGTGGCTGACCTGCGGCTGCTCATCTTCC 112

QY 856 GCTGCGCGCGCGCTCAGCTGCTGCTTGGCTCGGAGGACCTTCGCTGCCCGCGGC 915
    |||||||
Db 111 GCTGCGCGCGCTCAGCTGCTGCTTGGCTCGGAGGACCTTCGCTGCCCGCGGC 52

QY 916 CTCACCGACCGACGCGCGGTGAGAGTGGCTGCTGGAAGAGAGAACTGA 966
    |||||||
Db 51 CTCACCGACCGACGCGCGGTGAGAGTGGCTGCTGGAAGAGAGAACTGA 1

RESULT 8
AB032980/c 6548 bp mRNA linear PRI 10-MAY-2002
LOCUS Homo sapiens mRNA for KIAA1154 protein, partial cds.
DEFINITION AB032980
ACCESSION AB032980
VERSION AB032980.2 GI:20521781
KEYWORDS Homo sapiens adult male brain cDNA to mRNA, clone_lib:phluescriptII
SOURCE SK plus clone:hn03679s1.
ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hirosewa,M., Nagase,T., Ishikawa,K., Kikuno,R., Nomura,N. and
          Ohara,O.
TITLE Characterization of cDNA clones selected by the Genemark analysis
JOURNAL DNA Res. 6 (5), 329-336 (1999)
MEDLINE 20039618
PUBMED 10574461
REFERENCE 2 (bases 1 to 6548)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute,
          Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba
          292-0812, Japan (E-mail:cdna@infokazusa.or.jp.
          URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,
          Fax:+81-438-52-3914)
COMMENT On May 9, 2002 this sequence version replaced gi:6330108.
FEATURES
SOURCE location/Qualifiers
          1..6548

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hn03679s1"
/sex="male"
/tissue-type="brain"
/clone_lib="phluescriptII SK plus"
/seq_stage="adult"
/note="This sequence was obtained by subcloning of the DNA
fragments derived from two cDNA clones (1 - 1251 was
derived from a RT-PCR product (kidney) and 1252 - 6548 was
derived from hn03679)."
1..6548
/gene="KIAA1154"
<1..1431
/gene="KIAA1154"
/note="Start codon is not identified."
/codon_start=1
/product="KIAA1154 protein"
/protein_id="BA86468.2"
/db_xref="GI:20521782"
/translation="MSSSSARSSSLSPVYKSVLYRNGDPFYAGRPVITHEKVSF
EFLKEVTGVQAPFGAVRNITPTGHRRLKLDQISGNYAGGQAPRKLVYDI
GEIKRPMEEVNTVEKPVILSRINVSARFKPLDPECTIFLIANGDLINASRLIIP
KTLNOMDHVLOMYTEKLTLSGAVHRLYTEGLVESGAELENGQFVYAVGRDFKL
PYGLLFDKSTMRPFGOKASLPIYIGSKRSKSGKSDRSKSTVSGSDNSPOPLR
KGRKEVDSEKIFKLNKONVLEKINSOELIIPNSDEIFFAGERSETRGAAFOEDRQ
VEYVVDORPAEIVDEEDERKANDAKQKEDFSGMNGLDEEGREXTDAPQVEYL
DHSFOCARPAVNGTDEEKEELQOVNNEQLVLDKERKSGSGODEADVDQPR
PRPEVKITSPSENNNOONKDYAVA"
BASE COUNT 2080 a 1124 c 1377 g 1967 t
ORIGIN

Query Match      21.2%  Score 293;  DB 9;  Length 6548;
Best Local Similarity 100.0%  Pred. No. 3.8e-164;
Matches 293;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

556 TTGAGTTCTTGAAGGCTTCTGGCCCTCCAGCAGCAAGTAATTCGCCCGCTCGATCTGG 615
    |||||||
Db 293 TTGAGTTCTTGAAGGCTTCTGGCCCTCCAGCAGCAAGTAATTCGCCCGCTCGATCTGG 234

QY 616 TCTAGCTTCGGATTCCGTGGCCAGTCCGCGGGGTGTAGATTCTCTGACGCGCCCAAG 675
    |||||||
Db 233 TCTAGCTTCGGATTCCGTGGCCAGTCCGCGGGGTGTAGATTCTCTGACGCGCCCAAG 174

QY 676 GGTGCTTGAAGCGCGCGGTCTACCTCTTCAAGAAAGATTGGAAGCTGACACCTTCTTC 735
    |||||||
Db 173 GGTGCTTGAAGCGCGCGGTCTACCTCTTCAAGAAAGATTGGAAGCTGACACCTTCTTC 114

QY 736 TCATGATGACGACGCGCGCGCGGTGAGAAAGGCGTCCCGTTGCGGTACAAAGACAG 795
    |||||||
Db 113 TCATGATGACGACGCGCGCGCGGTGAGAAAGGCGTCCCGTTGCGGTACAAAGACAG 54

QY 796 CTCCTTACGACGCGCGGTGAGAGTGGCTGACCTGCGGCTGCTCCGCTCAT 848
    |||||||
Db 53 CTCCTTACGACGCGCGGTGAGAGTGGCTGACCTGCGGCTGCTCCGCTCAT 1

RESULT 9
LMFP1046/c 221641 bp DNA linear INV 05-SEP-2001
LOCUS Leishmania major Friedlin chromosome 13 PAC P1046.
DEFINITION AL359683
ACCESSION AL359683
VERSION AL359683.4 GI:15487186
KEYWORDS C3HC4 zinc-finger protein; DEAD-box helicase; DNA polymerase I,
          thermostable; exonuclease; flagellar radial spoke protein;
          guanosine diphosphatase; long-chain-fatty-acid-coa ligase;
          meak-related kinesin; mitogen-activated protein kinase; MORN-domain
          protein; n-acetyltransferase subunit; nucleolar protein involved in
          pre-rRNA; Pdz-domain protein; protein kinase a regulatory subunit;
          ras-related protein; surfeit 1-related protein; thermostable
          carboxypeptidase I; trab5b; tubulin alpha chain; Y429a.3 protein.
          Leishmania major.
SOURCE Leishmania major
ORGANISM

```

REFERENCE 1 (bases 1 to 221641)  
AUTHORS Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.  
TITLE A physical map of the Leishmania major Friedlin genome  
JOURNAL Genome Res. 8 (2), 135-145 (1998)  
MEDLINE 98146435  
PUBMED 9477341  
2 (bases 1 to 221641)  
AUTHORS Aert,R., Robben,J., Volckaert,G., Ivens,A.C., Quail,M., Rajadaram,M.A. and Barrell,B.G.  
TITLE Direct Submission  
JOURNAL Submitted (04-SEP-2001) European Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk) and Katholieke Universiteit Leuven, Faculty of Agricultural and Applied Biological Sciences, Department of Animal Production, Laboratory of Gene Technology, Willem de Croylaan 42  
On Sep 6, 2001 this sequence version replaced gi:13159932.  
see <http://www.ebi.ac.uk/parasites/leish.html>

## NOTES:

Details of leishmania sequencing at the Sanger Centre are available on the World Wide Web.  
see [http://www.sanger.ac.uk/Projects/L\\_major/](http://www.sanger.ac.uk/Projects/L_major/)  
CDS are numbered using the following system eg P1046.01. P1046 (PAC name), .01 (first CDS)  
To make the PAC library Leishmania major Friedlin DNA was partially digested with NotI prior to cloning into the NotI site of the PAC vector PCIRAC2N.  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
Gene prediction is done using:

(1) The Frameplot program of Blbb et al.,  
Gene 30:157-166(1984) as implemented  
at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. (2)  
codon preference based on the codon usage table for Leishmania at <http://www.kazusa.or.jp/codon/>  
(3)

the Hexamer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C. elegans Genome Sequencing Project. The program calculates the log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leishmania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg) which is preceded by a stretch of pyrimidines or part of a Kozak sequence. If this cannot be identified we choose the most upstream initiation codon. Transmembrane domains were predicted as implemented at the TMHMM server: <http://www.cbs.dtu.dk/services/TMHMM-1.0/>  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmids L2802 (AL359774) and L6293 (AL359777) are contained in PAC P1046. PAC P1046 is overlapped by cosmid L7913 (AL603794).

## FEATURES

source 1..221641  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="13"  
/clone="PAC\_P1046"  
complement(1..1227)  
/gene="P1046.01"

## CDS

complement(<1..1227)  
/gene="P1046.01"  
/note="P1046.01, predicted protein, len = 408 aa, unknown; predicted pi = 10.6399; contains no predicted TM helices; some similarity to APE\_CANFA, apolipoprotein e (305 aa, Canis familiaris, EMBL:); Fasta scores: E(1):0.59, 27.879% identity (29,677% ungapped) in 165 aa overlap, (aa 12-173 of P1046.01, aa 84-241 of APE\_CANFA)"  
/codon\_start=1  
/product="hypothetical predicted protein P1046.01, unknown function"  
/protein\_id="CAC37106.2"  
/db\_xref="GI:15487187"  
/db\_xref="SPRREMBL:O9BHY0"  
/translation="MSHICAGGSPSPILPLGATAGTSDVNCVPSQSDPHDAV KYERILRGAVYQKRLREARQKRKSEALASITKPPSELQRLFPVPSLASA ETRAHRAERQSALEKRAATAYCHPAIIPSEMASSRRRNMWQLVGGVLMRH KRVAGLEERKQOEVARFSEFPTISARATMOSRPEVRLSGTADCTDITFTD TNRSSGSEKTKASMATYORLTQDAVTRRAHDOEKARVSKNEFPCPDVSGD IASKRDTTREERLRKALDVAADVPOETPAPRINTSTREFRPVLSARAIIWRR SERLHAIVAVRACVEEMRECTFHRTRGRLTALSQAARATADSYTGPLASINSABEL LRVVDLEASAPPESECR"  
complement(1293..1311)  
/note="poly-pyrimidine tract"  
complement(1729..1743)  
/note="poly-pyrimidine tract"  
complement(1757..1771)  
/note="poly-pyrimidine tract"  
complement(1820..1837)  
/note="poly-pyrimidine tract"  
complement(2790..2805)  
/note="poly-pyrimidine tract"  
complement(2881..2911)  
/note="poly-pyrimidine tract"  
complement(3457..7746)  
/gene="P1046.02"  
complement(3457..7746)  
/note="P1046.02, predicted protein, len = 1428 aa, possibly DNA polymerase I, thermostable; predicted pi = 8.3748; contains Pfam match to entry PF00476 DNA\_pol\_A; contains match to PROSITE PS00447 DNA polymerase Family A signature; contains Pfam match to entry PF00929 Exonuclease; contains no predicted TM helices; reasonable similarity to DPO1\_THRAO, DNA polymerase I, thermostable (BC 2.7.7.7) (832 aa, Thermus aquaticus, EMBL: D32013, BA066775); Fasta scores: E(1):6.9e-07, 28.045% identity (32,143% ungapped) in 353 aa overlap, (aa 1021-1366 of P1046.02, aa 515-829 of DPO1\_THRAO)"  
/codon\_start=1  
/product="possible DNA polymerase I, thermostable"  
/protein\_id="CAC37107.1"  
/db\_xref="GI:13751093"  
/db\_xref="SPRREMBL:O9BHY9"  
/translation="MFRIGSSMPRALAQLPLRSSCTVQSVSLTAQAALARGAPSVDA AYAEEFRVPIVDAGVVRSTQDEYTNIRLAENQSDVAQNSPKSGSDSVATFL VYHPTQATFOEFLISAADIVCSVTRMNASVQSDADELMHIGESPOLALPVE FTERKRILOEPLOKTLKESGLOVATVDHLELSVYCSNPPSPRAITQDAVSGV SPMPRTMGOAMATRTTAAOCAMSDPQVSTYTRDPRFVYLIHLVRAAGRLANS LMDPVSRRFRSKMGLSDVLCILDRSSQTLVPTTRDRALRCRAVAVKAGHDC VVAVSVESELRGCGYQSDLMRWGAKMTAFSLDADLFGITYLAGNTLRKYORKL SQSHCDNLISLNSDSEGATPQNLNMGVCTLPKRVAAEKELRHVDKFTTIVYS THTVTAQVDSNVADINHLASACITDHANRTVEPAVITKRGQFCLPSLDGDLVTH DAKSLVLTLSGAELOKFTKRGGRVWCYTLAEYLEAQCSTSLDHLARHCVOL PPSRVGTPNDLPRALQRLHLHAARVAVYFVQKRALEQCYLSLHMDLSLEL MTSIEKAGIHDAEENAROTASLSKASVLDAMAEVYSEVPVMRTYFDWASLQOQ HAYFEGGTTISLHGOAVARDPTPLMTSNVHLCHRGYAFHLMWGLHQRAAACALPTS GGLPSRIHQHIEVQOSOKLTQRYVEFDETGLPSTDAIVEAMFPIENSTFTPL VNPORRTPRTGVGHIHTANAVRDAVPTDVVAKSIGVYLRDKASYNPDETLVGHN VPSVDEPMRLRALEHAPRCODGLTFQSLALNHAHSELEWRRRTKRTNPLMDAL ASSLRSLRITLNVPRBGDLIRATDTRKALMYVLINTLGVADKRAVAKORDTLTEAA NCFLSPSTIGCFVPAERONRIVKVRPGVAADVVRNAKILASQKSEVSEVYASLHA HGVKPAGLLQRLDLRHTSTFLQSPSTGRLALILHRDGHVHIDMTATTTTSRTVAS

```

PSCQNIIPKDKSSVRRLVFRSGKRCVEVDYSOLEIVLAILCNDANLTJNDLSGV
DEHYKRAAFESGLPYDEIYOGYKRNVPKYLKRTAKOFSPQRLYGAGYPLHKTTGI
PYKDLKASIORENEEYPIAGOFHRIIRSVLRPNNGLPSPFLAEMPTGLRSLRTGD
VYVNLNLPITNPIYOGAGELIOMMGLRTRHFRKROFYPDRAPFLINFEVDSVMDOCHV
DYLRREVPRDTCRIILGSHYHYKVPFGVATISVPLOYASACGCDMCMESITKGDYMFV
SKORTRSAEVDPDFDLTTAKNSFVMEESVASEEGSGETATSEATE"

misc_feature
    complement(3646..4806)
    /gene="P1046.02"
    /note="pfam match to entry PF00476 DNA_pol_A, DNA
    polymerase family A, score -62.00, E-value 9.6e-08"

misc_feature
    complement(4153..4212)
    /gene="P1046.02"
    /note="match to PROSITE P50447 DNA polymerase family A
    signature"
    complement(4876..5457)
    /gene="P1046.02"
    /note="pfam match to entry PF00929 Exonuclease,
    Exonuclease, score 63.00, E-value 6.5e-15"

repeat_region
    /note="(TGC)4"
    5699..5710
    /note="poly-pyrimidine tract"
    8143..8160
    /note="poly-pyrimidine tract"
    8288..8303
    /note="poly-pyrimidine tract"
    complement(8822..8842)
    /note="poly-pyrimidine tract"
    complement(8882..8898)
    /note="poly-pyrimidine tract"
    complement(9244..9263)
    /note="poly-pyrimidine tract"
    complement(9311..9341)
    /note="poly-pyrimidine tract"
    complement(9357..9373)
    /note="poly-pyrimidine tract"
    complement(9472..9487)
    /note="poly-pyrimidine tract"
    complement(9851..9866)
    /note="poly-pyrimidine tract"
    complement(10073..10108)
    /note="poly-pyrimidine tract"

Query Match
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1216 AGAGGAGGACGACACACACACACACA 1246
|||||
Db 187446 AGAGGAGGACGACACACACACACACA 187416

RESULT 10
AC024857/c 38967 bp DNA linear INV 19-APR-2002
DEFINITION Caenorhabditis elegans cosmid Y71G12A, complete sequence.
ACCESSION AC024857
VERSION AC024857.2 GI:20198913
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Rhabditioidea: Rhabditidae;
Rhabditioidea: Rhabditidae; Peloderinae: Caenorhabditis
1 (bases 1 to 38967)
Waterston,R.
REFERENCE genome sequence of the nematode C. elegans: a platform for
AUTHORS investigating biology. The C. elegans Sequencing Consortium
TITLE Science 282 (5396), 2012-2018 (1998)
JOURNAL 99069613
MEDLINE 9851916
PUBMED 2 (bases 1 to 38967)
REFERENCE Harmon,G., Lamar,B., Du,H. and Wohlmann,P.
AUTHORS The sequence of C. elegans cosmid Y71G12A
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 38967)

```

```

AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 38967)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 38967)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 6 (bases 1 to 38967)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 7 (bases 1 to 38967)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
REFERENCE 8 (bases 1 to 38967)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On Apr 19, 2002 this sequence version replaced gi:7140420.
Submitted by:
    Genome Sequencing Center
    Department of Genetics, Washington University
    St. Louis, MO 63110, USA, and
    Sanger Centre, Hinxton Hall
    Cambridge CB10 1RQ, England
    email: tw@nematoe.wustl.edu and jesus@sanger.ac.uk

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m33 subclone.

For a graphical representation of this cosmid sequence and its analysis see:  
{www.wormbase.org/db/seq/sequence?name=Y71G12A:class=Sequence}

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is M01D7, 3500 bp overlap; the 3' cosmid is Y51F10, 800 bp overlap.

#### NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFINDER (P. Green and V. Hillier, personal communication), the large scale EST projects of Yui Kohara ([http://www.ddb.jig.ac.jp/c-elegans/html/CE\\_INDEX.html](http://www.ddb.jig.ac.jp/c-elegans/html/CE_INDEX.html)) and The C. elegans ORFome cloning project (<http://wormfdb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses



## COMMENT

University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA  
On Aug 9, 2001 this sequence version replaced gi:2105482.  
Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: twenematode.wustl.edu and jesusanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this cosmid sequence and its analysis see:

[www.wormbase.org/db/seq/sequence?name=M01D7;class=Sequence]

## NEIGHBORING COSMID INFORMATION

The 5' cosmid is C53H9, 3300 bp overlap; the 3' cosmid is Y71G12A, 3500 bp overlap. Actual start of this cosmid is at base position 3297 of M01D7; actual end is at 41179 of M01D7.

## NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujl Kohara (http://www.ddbj.nig.ac.jp/c-elegans/hum1/CE\_INDEX.html) and The C. elegans ORFome cloning project (http://wormfdb.dcfcl.harvard.edu/), similarity to other proteins from Blastx analyses (http://Blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

## FEATURES

## source

```
1.41179
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="M01D7"
52..5004
/gene="egl-30"
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=M01D7;?a:Class=Sequence]"
join(52..169,2275..2402,2896..3113,3180..3308,3367..3496,
4211..4364,4435..4542,4922..5004)
/gene="egl-30"
```

## CDS

## gene

```
/note="C. elegans heterotrimeric G protein (EGL-30)
(GB:U56864); contains similarity to Pfam domain: PF00503
(G-alpha). Score=761.7, E-value=1e-225, N=1; coded for by
the following C. elegans cDNAs: YK29G12.5, YK32962.5,
YK277B4.5, YK10466.5, YK17669.5, YK36166.5, YK28966.5,
YK343912.5, YK413B9.5, YK27219.5, YK46266.5, YK461B9.5,
YK22611.5, YK215A10.5, YK20366.5, YK343912.3, YK22812.5,
YK52969.5, YK551d12.5, YK54466.5, YK630H4.5, YK536d9.5,
YK587A10.5, YK681A11.5, YK622A12.3, YK65466.3, YK68168.3,
YK892A08.3, YK1343A05.3, YK551B6.5, YK1247E07.5,
```

```
YK892A08.5, YK1343A05.5, U56864"
/codon_start=1
/product="C. elegans EGL-30 protein (corresponding
sequence M01D7.7a)"
/db_xref="GI:2105489"
```

## gene

```
/translation="MACCSSEAREQRKINQEIETKQLRDRKRNARRELKLLTGTS
GKSPFKOMRIHOGYSEEDKRAHILYONVPMALQSMRAMDTLDIRGNSEEL
OKRAAVYREVDPESTVTFEEPYYSIKELMEDSGIOCYDRRRYQITDSAKYLSL
RLAVPDLPTEDDILRVAPPTGIIIEPDLQIIRFMDVGQSRERKWHCFEN
VTSIMFLVALSEYDOVLVECDNEMESKALFTITITFPWTSVILFLNKKDLLE
EKILYSHLADYFPEYDGPDPPIAREFILKMFVDLNDADKLIYSHTCATDTENIR
FVPAAYVDTLIQHNLKEYNLV"
```

## CDS

```
1097..1886
/gene="M01D7.8"
```

```
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=M01D7.8;class=Sequence]"
join(1097..1233,1814..1886)
```

## CDS

```
/gene="M01D7.8"
/codon_start=1
/product="Hypothetical protein M01D7.8"
```

```
/protein_id="AA015594.1"
/db_xref="GI:20198866"
/translation="MEKREEREGSGSHNLISANQPTNQAPPTTGARKSPERK
WGFHRELIYLDQKTRKRRVNLN"
join(2313..2402,2896..3113,3180..3308,3367..3496,
4211..4364,4435..4542,4922..5004)
```

```
/gene="egl-30"
/note="C. elegans heterotrimeric G protein (EGL-30);
alternatively-spliced form; coded for by the following C.
elegans cDNAs: YK51433.5"
```

```
/codon_start=1
/product="C. elegans EGL-30 protein (corresponding
sequence M01D7.7b)"
```

```
/protein_id="AA015593.1"
/db_xref="GI:20198865"
```

## misc\_feature

```
/translation="MRIHOGYSEERKRAHILYONVPMALQSMRAMDTLDIRK
NESELDKAAVREVDPESTVTFEEPYYSIKELMEDSGIOCYDRRRYQITDSAK
YLSLDRLRAVPYLPEDDILRVAPPTGIIIEPDLQIIRFMDVGQSRERKWHCFEN
IHCFENTSIMFLVALSEYDOVLVECDNEMESKALFTITITFPWTSVILFLNKKDLLE
KILYSHLADYFPEYDGPDPPIAREFILKMFVDLNDADKLIYSHTCATDTENIR
FVPAAYVDTLIQHNLKEYNLV"
```

## gene

```
6560..6561
/note="5' trans-splice site; see YK589H7.5"
```

```
6576..7202
/gene="M01D7.6"
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=M01D7.6;class=Sequence]"
join(6576..6714,6789..7039,7092..7202)
```

## CDS

```
/gene="M01D7.6"
/note="similar to a short region of thymopletins; coded
for by the following C. elegans cDNAs: YK79H4.5,
```

```
YK13610.5, YK113E10.3, YK19H4.3, YK442H6.3, YK142H6.5,
YK196B5.5, YK258G11.3, YK258H1.5, YK466F7.5, YK589H7.5,
YK686D10.5, YK669G2.3, YK668D10.3, YK589H7.3"
/codon_start=1
/product="Hypothetical protein M01D7.6"
```

```
/protein_id="AA058065.1"
/db_xref="GI:2105483"
/translation="MDVSOI/TDAELRSLKSHGVNPIVATPRKLYEKRLISDGS
ITNOSNLNDSQFNEDSLITSSPKRSPPQRYOVNSATATATSPSDSDDECSERK
YLTEEEAADRASARKQSNKGGFLSGTITFTLTVFAVFAFLIENADQLKVAET
NPBDIT"
```

## gene

```
7523..7901
/gene="nlp-12"
```

## CDS

```
/note="for a graphical representation of this gene see:
[www.wormbase.org/db/seq/sequence?name=M01D7.5;class=Sequence]"
join(7523..7577,7657..7715,7773..7901)
```

```
/gene="nlp-12"
/note="C. elegans putative neuropeptide preproprotein
```

(NLP-12): coded for by the following C. elegans cDNAs:  
yk162h11.3, yk162h11.5, yk276h6.3, yk276h6.5, yk1188b11.3,  
yk550c1.5"  
/codon\_start=1  
/product="C. elegans NLP-12 protein (corresponding  
sequence M01D7.5)"  
/protein\_id="AB58066.1"  
/db\_xref="GI:2105484"  
/translation="MLRHSCALMLILIVFVFATQSPFPDRODRYRPLQGRKDG  
YRPLQGRKDYRPLQGRKSSSSSGPVLEPIWMQ"  
9337, 14518  
/gene="M01D7.4"  
/note="for a graphical representation of this gene see:

Query Match 2.1%; Score 29; DB 3; Length 41179;  
Best local Similarity 100.0%; Pred. No. 4.3e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1218 AGGAGACGACACACACACACACACA 1246  
DB 38199 AGGAGACGACACACACACACACACA 38171

RESULT 12  
AC100574 49743 bp DNA linear HTG 22-NOV-2001  
LOCUS Mus musculus clone RP23-155010, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC100574  
AC100574 1 GI:17047940  
VERSION HTG; HTGS\_PHASE0.  
KEYWORDS  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 49743)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
JOURNAL Mus musculus, clone RP23-155010  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 49743)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barra,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campoliano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
Cook,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kellis,C., Lacroque,K.,  
Lamatzes,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,  
Menus,L., Mihova,T., Mienna,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Topham,K., Travers,M., Travers,N., Trigilio,D., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information

Center project name: L15715  
Center clone name: 155\_O\_10

\* NOTE: This record contains 62 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 681: contig of 681 bp in length  
682 781: gap of 100 bp  
782 1494: contig of 713 bp in length  
1495 1594: gap of 100 bp  
1595 2314: contig of 720 bp in length  
2315 2414: gap of 100 bp  
2415 3097: contig of 683 bp in length  
3098 3197: gap of 100 bp  
3198 3901: contig of 704 bp in length  
3902 4001: gap of 100 bp  
4002 4679: contig of 678 bp in length  
4680 4779: gap of 100 bp  
4780 5453: contig of 674 bp in length  
5454 5553: gap of 100 bp  
5554 6250: contig of 697 bp in length  
6251 6350: gap of 100 bp  
6351 7065: contig of 715 bp in length  
7066 7165: gap of 100 bp  
7166 7905: contig of 740 bp in length  
7906 8005: gap of 100 bp  
8006 8710: contig of 705 bp in length  
8711 8810: gap of 100 bp  
8811 9511: contig of 701 bp in length  
9512 9611: gap of 100 bp  
9612 10305: contig of 694 bp in length  
10306 10405: gap of 100 bp  
10406 11114: contig of 709 bp in length  
11115 11214: gap of 100 bp  
11215 11952: contig of 738 bp in length  
11953 12052: gap of 100 bp  
12053 12712: contig of 660 bp in length  
12713 12812: gap of 100 bp  
12813 13529: contig of 717 bp in length  
13530 13629: gap of 100 bp  
13630 14351: contig of 722 bp in length  
14352 14451: gap of 100 bp  
14452 15153: contig of 702 bp in length  
15154 15253: gap of 100 bp  
15254 15973: contig of 720 bp in length  
15974 16073: gap of 100 bp  
16074 16774: contig of 701 bp in length  
16775 16874: gap of 100 bp  
16875 17545: contig of 671 bp in length  
17546 17645: gap of 100 bp  
17646 18341: contig of 696 bp in length  
18342 18441: gap of 100 bp  
18442 19166: contig of 725 bp in length  
19167 19266: gap of 100 bp  
19267 20005: contig of 739 bp in length  
20006 20105: gap of 100 bp  
20106 20827: contig of 722 bp in length  
20828 20927: gap of 100 bp  
20928 21622: contig of 695 bp in length  
21623 21722: gap of 100 bp  
21723 22436: contig of 714 bp in length  
22437 22536: gap of 100 bp  
22537 23336: contig of 700 bp in length  
23337 23337: gap of 100 bp  
23337 24029: contig of 693 bp in length

```

* 24030 24129: gap of 100 bp
* 24130 24816: contig of 687 bp in length
* 24817 24916: gap of 100 bp
* 24917 25610: contig of 694 bp in length
* 25611 25710: gap of 100 bp
* 25711 26447: contig of 737 bp in length
* 26448 26547: gap of 100 bp
* 26548 27268: contig of 721 bp in length
* 27269 27368: gap of 100 bp
* 27369 28071: contig of 703 bp in length
* 28072 28171: gap of 100 bp
* 28172 28871: contig of 700 bp in length
* 28872 28971: gap of 100 bp
* 28972 29671: contig of 700 bp in length
* 29672 29771: gap of 100 bp
* 29772 30493: contig of 722 bp in length
* 30494 30593: gap of 100 bp
* 30594 31314: contig of 721 bp in length
* 31315 31414: gap of 100 bp
* 31415 32111: contig of 697 bp in length
* 32112 32211: gap of 100 bp
* 32212 32905: contig of 694 bp in length
* 32906 33005: gap of 100 bp
* 33006 33703: contig of 698 bp in length
* 33704 33803: gap of 100 bp
* 33804 34463: contig of 660 bp in length
* 34464 34563: gap of 100 bp
* 34564 35269: contig of 706 bp in length
* 35270 35369: gap of 100 bp
* 35370 36053: contig of 684 bp in length
* 36054 36153: gap of 100 bp
* 36154 36884: contig of 731 bp in length
* 36885 36984: gap of 100 bp
* 36985 37704: contig of 720 bp in length
* 37705 37804: gap of 100 bp
* 37805 38527: contig of 723 bp in length
* 38528 38627: gap of 100 bp
* 38628 39346: contig of 719 bp in length
* 39347 39446: gap of 100 bp
* 39447 40161: contig of 715 bp in length
* 40162 40261: gap of 100 bp
* 40262 40927: contig of 666 bp in length
* 40928 41027: gap of 100 bp
* 41028 41708: contig of 681 bp in length
* 41709 41808: gap of 100 bp
* 41809 42495: contig of 687 bp in length
* 42496 42595: gap of 100 bp
* 42596 43278: contig of 683 bp in length
* 43279 43378: gap of 100 bp
* 43379 44113: contig of 735 bp in length
* 44114 44213: gap of 100 bp
* 44214 44920: contig of 707 bp in length
* 44921 45020: gap of 100 bp
* 45021 45694: contig of 674 bp in length
* 45695 45794: gap of 100 bp
* 45795 46496: contig of 702 bp in length
* 46497 46596: gap of 100 bp
* 46597 47305: contig of 709 bp in length
* 47306 47405: gap of 100 bp
* 47406 48090: contig of 685 bp in length
* 48091 48190: gap of 100 bp
* 48191 48909: contig of 719 bp in length
* 48910 49009: gap of 100 bp
* 49010 49743: contig of 734 bp in length.

```

## FEATURES

SOURCE

1..49743

/organism="Mus musculus"

/db\_xref="taxon:10090"

```

BASE COUNT 12369 a 9149 c 8642 g 13293 t 6290 others
ORIGIN

```

```

Query Match          2.1%; Score 29; DB 2; Length 49743;
Best Local Similarity 100.0%; Pred. No. 4.4e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1227 CACACACACACACACACAAATATG 1255
|||||
DB 31047 CACACACACACACACACAAATATG 31075

```

```

RESULT 13
AC079517_3/c
WPCOMMENT
Sequence split into 5 fragments

```

Fragment Name	Begin	End	LOCUS	AC079517	Accession	AC079517
AC079517_0	1	110000				
AC079517_1	100001	210000				
AC079517_2	200001	310000				
AC079517_3	300001	410000				
AC079517_4	400001	497254				

Continuation (4 of 5) of AC079517 from base 300001 (AC079517 Mus musculus clone RP23-

```

Query Match          2.1%; Score 29; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1224 ACCGACACACACACACACAAATATG 1252
|||||
DB 82894 ACCGACACACACACACACAAATATG 82866

```

```

RESULT 14
AC097411/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-142N16, *** SEQUENCING IN PROGRESS
***, 69 unsorted pieces.
ACCESSION
AC097411.4 GI:21728566
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 119318)
REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aldisrooks,S.L., Amaralunge,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbarta,J., Benton,J., Blinage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzner,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbad,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogum,M., Okwuon,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

```

Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,  
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swalek, A., Taber, P., Tameris, A., Tameris, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, F., Zhou, J., Zorrilla, S., Nelson, D.,  
Meinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 119318)  
Worley, K.C.

Direct Submission  
Submitted (18-OCT-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 119318)  
Worley, K.C.

Direct Submission  
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gi:17973555.

----- Genome Center  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-142N16  
Center clone name: CH230-142N16  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 58551 bases at least Q40  
Consensus quality: 62883 bases at least Q30  
Consensus quality: 66623 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 69 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1014: contig of 1014 bp in length  
\* 1015 1114: gap of unknown length  
\* 1115 2122: contig of 1008 bp in length  
\* 2123 2222: gap of unknown length  
\* 2223 3409: contig of 1187 bp in length  
\* 3410 3509: gap of unknown length  
\* 3510 4530: contig of 1021 bp in length  
\* 4531 4630: gap of unknown length  
\* 4631 5933: contig of 1303 bp in length  
\* 5934 6033: gap of unknown length  
\* 6034 7040: contig of 1007 bp in length  
\* 7041 7140: gap of unknown length  
\* 7141 8321: contig of 1181 bp in length  
\* 8322 8421: gap of unknown length  
\* 8422 9774: contig of 1353 bp in length  
\* 9775 9874: gap of unknown length  
\* 9875 11065: contig of 1191 bp in length  
\* 11066 11165: gap of unknown length  
\* 11166 12218: contig of 1053 bp in length  
\* 12219 12318: gap of unknown length  
\* 12319 13334: contig of 1016 bp in length

13335 13434: gap of unknown length  
\* 13436 14534: contig of 1100 bp in length  
\* 14535 14634: gap of unknown length  
\* 14635 15998: contig of 1364 bp in length  
\* 15999 16098: gap of unknown length  
\* 16099 17453: contig of 1355 bp in length  
\* 17454 17553: gap of unknown length  
\* 17554 18675: contig of 1122 bp in length  
\* 18676 18776: gap of unknown length  
\* 18777 19889: contig of 1114 bp in length  
\* 19890 19989: gap of unknown length  
\* 19990 21020: contig of 1031 bp in length  
\* 21021 21120: gap of unknown length  
\* 21121 22170: contig of 1050 bp in length  
\* 22171 23508: contig of 1238 bp in length  
\* 23509 23608: gap of unknown length  
\* 23609 24674: contig of 1066 bp in length  
\* 24675 26176: gap of unknown length  
\* 26177 26276: gap of unknown length  
\* 26277 27394: contig of 1118 bp in length  
\* 27395 27494: gap of unknown length  
\* 27495 28707: contig of 1213 bp in length  
\* 28708 28807: gap of unknown length  
\* 28808 30449: contig of 1642 bp in length  
\* 30450 30549: gap of unknown length  
\* 30550 31742: contig of 1193 bp in length  
\* 31743 31842: gap of unknown length  
\* 31843 33184: contig of 1342 bp in length  
\* 33185 33284: gap of unknown length  
\* 33285 34308: contig of 1024 bp in length  
\* 34309 34408: gap of unknown length  
\* 34409 36022: contig of 1614 bp in length  
\* 36023 36122: gap of unknown length  
\* 36123 38458: contig of 2336 bp in length  
\* 38459 38558: gap of unknown length  
\* 38559 39841: contig of 1283 bp in length  
\* 39842 39941: gap of unknown length  
\* 39942 41425: contig of 1484 bp in length  
\* 41426 41525: gap of unknown length  
\* 41526 42987: contig of 1462 bp in length  
\* 42988 43087: gap of unknown length  
\* 43088 44582: contig of 1495 bp in length  
\* 44583 44682: gap of unknown length  
\* 44683 46773: contig of 2091 bp in length  
\* 46774 46873: gap of unknown length  
\* 46874 47917: contig of 1044 bp in length  
\* 47918 48017: gap of unknown length  
\* 48018 49934: contig of 1917 bp in length  
\* 49935 50034: gap of unknown length  
\* 50035 51300: contig of 1266 bp in length  
\* 51301 51400: gap of unknown length  
\* 51401 52566: contig of 1166 bp in length  
\* 52567 52666: gap of unknown length  
\* 52667 53745: contig of 1079 bp in length  
\* 53746 53845: gap of unknown length  
\* 53846 55131: contig of 1286 bp in length  
\* 55132 55231: gap of unknown length  
\* 55232 56364: contig of 1133 bp in length  
\* 56365 56464: gap of unknown length  
\* 56465 57510: contig of 1046 bp in length  
\* 57511 57610: gap of unknown length  
\* 57611 60028: contig of 2418 bp in length  
\* 60029 60128: gap of unknown length  
\* 60129 61260: contig of 1132 bp in length  
\* 61261 61360: gap of unknown length  
\* 61361 63338: contig of 1978 bp in length  
\* 63339 63438: gap of unknown length  
\* 63439 64994: contig of 1556 bp in length  
\* 64995 65094: gap of unknown length  
\* 65095 66620: contig of 1526 bp in length  
\* 66621 66720: gap of unknown length



```
*      66721      68397: contig of 1677 bp in length
*      68398      68497: gap of unknown length
*      68498      70503: contig of 2106 bp in length
*      70604      70703: gap of unknown length
*      70704      72233: contig of 1530 bp in length
*      72234      72333: gap of unknown length
*      72334      73576: contig of 1243 bp in length
*      73577      73676: gap of unknown length
*      73677      74814: contig of 1138 bp in length
*      74815      74914: gap of unknown length
*      74915      77154: contig of 2240 bp in length
*      77155      77254: gap of unknown length
*      77255      78472: contig of 1218 bp in length
```

Query Match 2.1%; Score 29; DB 2; Length 119318;  
Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 ACGCACACACACACACACAATATG 1252  
|||||  
DB 2880 ACGCACACACACACACACAATATG 2852

RESULT 15  
AC020895/c 139846 bp DNA linear PRI 20-JUL-2001  
LOCUS Homo sapiens chromosome 19 clone CTC-312010, complete sequence.  
AC020895  
AC020895.8 GI:14971187  
VERSION HTG.  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 139846)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission

REFERENCE  
AUTHORS 2 (bases 1 to 139846)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission

REFERENCE  
AUTHORS Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE  
AUTHORS 3 (bases 1 to 139846)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission

REFERENCE  
AUTHORS Submitted (26-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

REFERENCE  
AUTHORS 4 (bases 1 to 139846)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Direct Submission

REFERENCE  
AUTHORS Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

REFERENCE  
AUTHORS On Jul 20, 2001 this sequence version replaced gi:14550304.  
JOURNAL Draft Sequence Produced by DOE Joint Genome Institute

REFERENCE  
AUTHORS www.fgi.doe.gov  
JOURNAL Finishing Completed at Stanford Human Genome Center

REFERENCE  
AUTHORS www.shgc.stanford.edu  
JOURNAL Quality: Phrap Quality >=40 99.3% of Sequence;

REFERENCE  
AUTHORS Estimated Total Number of Errors is 0.5.  
JOURNAL STS Content:

SHGC-35395 G28602.

FEATURES  
Source Location/Qualifiers  
1..139846

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"

BASE COUNT 37798 a 30655 c 31600 g 39793 t  
ORIGIN

Query Match 2.1%; Score 29; DB 9; Length 139846;  
Best Local Similarity 100.0%; Pred. No. 4.9e-05;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 ACGCACACACACACACACAATATG 1252  
|||||  
DB 132576 ACGCACACACACACACACAATATG 132548

Search completed: January 18, 2003, 23:06:10  
JOB time : 4141 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 20:46:59 ; Search time 2211 Seconds  
(without alignments)  
10123.095 Million cell updates/sec

Title: US-09-674-593-1

Perfect score: 1382

Sequence: 1 cattaatcctaagcataaa.....taaagaacttaagtctctgg 1382

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	607	43.9	879	14	B0953400 AGENCOURT
2	582	42.1	845	14	B0948660 AGENCOURT
3	553	40.0	898	14	B0650550 AGENCOURT
4	501	36.3	963	14	B0650598 AGENCOURT
5	470	34.0	486	14	BM833169 K-EST0107
6	414	30.0	994	14	B0650059 AGENCOURT

c	7	342	24.7	984	14	BM923267
c	8	308	22.3	1045	12	BF749180
c	9	288	20.8	929	14	B0644671
c	10	288	20.8	939	14	B0644474
c	11	288	20.8	1019	14	BM927835
c	12	241	17.4	246	9	AA863443
c	13	196	14.2	357	10	AA863443
c	14	142	10.3	364	12	BF364659
c	15	142	10.3	835	12	BF750306
c	16	110	8.0	360	12	BF876266
c	17	86	6.2	923	12	BF51895
c	18	29	2.1	353	17	A2709419
c	19	29	2.1	585	17	BF334637
c	20	28	2.0	484	17	A0058290
c	21	27	2.0	385	17	A2656621
c	22	27	2.0	427	17	A0818586
c	23	27	2.0	439	12	A2012650
c	24	27	2.0	490	12	BE935673
c	25	27	2.0	633	17	A0633858
c	26	27	2.0	693	17	BH045618
c	27	27	2.0	849	17	A0748915
c	28	27	2.0	977	17	CNS020NS
c	29	26	1.9	89	9	AA500598
c	30	26	1.9	148	17	BH347122
c	31	26	1.9	201	17	A2741359
c	32	26	1.9	217	12	BG147613
c	33	26	1.9	235	12	BF950472
c	34	26	1.9	242	10	BH579813
c	35	26	1.9	256	17	B72773
c	36	26	1.9	268	10	BH083826
c	37	26	1.9	286	12	BF910381
c	38	26	1.9	310	17	A2080000
c	39	26	1.9	393	10	AA546635
c	40	26	1.9	398	12	BG072715
c	41	26	1.9	404	17	A2488659
c	42	26	1.9	410	12	BG085507
c	43	26	1.9	430	17	BH058773
c	44	26	1.9	459	17	B53308
c	45	26	1.9	462	17	A0224051

## ALIGNMENTS

RESULT 1  
B0953400 879 bp mRNA linear EST 21-AUG-2002  
LOCUS AGENCOURT\_8784199 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:6376296  
DEFINITION 5', mRNA sequence.  
ACCESSION B0953400  
VERSION B0953400.1 GI:22368878  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 879)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LCM2558 row: e column: 01  
High quality sequence stop: 690.  
Location/Qualifiers  
1..879

FEATURES  
source



```
Db 333 TGGCCAGTCCGGGGGTGTAGATGTTCTTGACGCCGCCAAAGGAGTGGCTTGAACGGCCGCC 274
Oy 694 GTACACTCTTTAGAGAACTTGAAGCTGGACACCTTTTCTCATGATGATGACAGCGG 753
Db 273 GTGACCTCTCTTCAAGAACTTGAAGCTGGACACCTTTTCTCATGATGATGACAGCGG 214
Oy 754 CGCCCCGGGTAGAAAGGGTCCCGTTGGGGTACACAGCAGCTCTTCACGAGCGGGCTGA 813
Db 213 CGCCCCGGGTAGAAAGGGTCCCGTTGGGGTACACAGCAGCTCTTCACGAGCGGGCTGA 154
Oy 814 GACAGGTGGCTGGACCTGGCGCTGCTCCATCTTCCCGCTGGCGCCGCCCTCAGC 873
Db 153 GACAGGTGGCTGGACCTGGCGCTGCTCCATCTTCCCGCTGGCGCCGCCCTCAGC 94
Oy 874 TCGCTGCTGGGTGGGAGGAGCAGCTCCGCTGTCCAGCGGGCTCAACGACGAGGCG 933
Db 93 TCGCTGCTGGGTGGGAGGAGCAGCTCCGCTGTCCAGCGGGCTCAACGACGAGGCG 34
Oy 934 CGGATCGCTCTTGAAACGAGAGAAACTGA 966
Db 33 CGGATCGCTCTTGAAACGAGAAACTGA 1

RESULT 3
BO650550 898 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8350342 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286493
DEFINITION 5', mRNA sequence.
ACCESSION BO650550
VERSION BO650550.1 GI:21774722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 898)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rt@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: L1CM2487 row: g column: 06
High quality sequence stop: 667.
Location/Qualifiers
1..898
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6286493"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Liver; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(C). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."
BASE COUNT 206 a 252 c 273 g 167 t
ORIGIN
```

```
Query Match 40.0%; Score 553; DB 14; Length 898;
Best Local Similarity 100.0%; Pred. No. 9,4e-272;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 396 CCCCCCACTGCAAGAGAGTTTGTGAGGGCATCAATCACACCGAGAAATGCACAGCCCT 455
Db 664 CCCCCCACTGCAAGAGAGTTTGTGAGGGCATCAATCACACCGAGAAATGCACAGCCCT 605
Oy 456 CAACCACTGAGTGTGGGGGTAGAGATCTGATTTCTTCAATATCAACCCCACTATA 515
Db 604 CAACCACTGAGTGTGGGGGTAGAGATCTGATTTCTTCAATATCAACCCCACTATA 545
Oy 516 GGGCACTTAATGGTGGGGGTGGGGAGACCGACTCACTTGAATTCTTGAAGGCTTC 575
Db 544 GGGCACTTAATGGTGGGGGTGGGGAGACCGACTCACTTGAATTCTTGAAGGCTTC 485
Oy 576 CTGGCTCCAGCAGCAATATGCCCCCTCTGAGATCTGTGTTAGCTTCCGATTCGGG 635
Db 484 CTGGCTCCAGCAGCAATATGCCCCCTCTGAGATCTGTGTTAGCTTCCGATTCGGG 425
Oy 636 GCCAGTCCGGGGGTGTAGATGTTCTTGACGGCCCCAAAGGTGCTGAACGCCCGCT 695
Db 424 GCCAGTCCGGGGGTGTAGATGTTCTTGACGGCCCCAAAGGTGCTGAACGCCCGCT 365
Oy 696 CACTCTCTTCAAGAACTTGAAGCTGACACCTTCTTCATGATGATGACAGCGGCG 755
Db 364 CACTCTCTTCAAGAACTTGAAGCTGACACCTTCTTCATGATGATGACAGCGGCG 305
Oy 756 CCCCCGTGAAGGGGTCCCGCTGGCGGTACACAGCAGGCTTTCACGACGGGCTGAGA 815
Db 304 CCCCCGTGAAGGGGTCCCGCTGGCGGTACACAGCAGGCTTTCACGACGGGCTGAGA 245
Oy 816 CAGTGTGCTGACAGCTGGCGCTGCTCCGCTCACTTCTCCCGTGGCGCGGCTCAGCTC 875
Db 244 CAGTGTGCTGACAGCTGGCGCTGCTCCGCTCACTTCTCCCGTGGCGCGGCTCAGCTC 185
Oy 876 GCTGCTTGGCTGGGAGGACCTCCGCTGCCACAGGCGCTCACCGCACCGAGGGCGG 935
Db 184 GCTGCTTGGCTGGGAGGACCTCCGCTGCCACAGGCGCTCACCGCACCGAGGGCGG 125
Oy 936 GGATCGCTCTCG 948
Db 124 GGATCGCTCTCG 112

RESULT 4
BO650598 963 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8207577 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283187
DEFINITION 5', mRNA sequence.
ACCESSION BO650598
VERSION BO650598.1 GI:21774770
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 963)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rt@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: L1CM2478 row: m column: 12
High quality sequence stop: 618.
Location/Qualifiers
1..963
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```

```

/clone="IMAGE:6283187"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT      171 a      350 c      244 g      197 t      1 others
ORIGIN
Query Match      36.3%; Score 501; DB 14; Length 963;
Best Local Similarity 100.0%; Pred. No. 4,1e-245;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 448 CAGCCCTCAACACAGTGGTGGGGGTAGGATCTGCATTTCTCATATCAACCCC 507
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 602 CAGCCCTCAACACAGTGGTGGGGGTAGGATCTGCATTTCTCATATCAACCCC 543
QY 508 ACACATATAGGACCTTAATGGGTGGCGGGTGGGAGACCGACTCATTTGATTCTTG 567
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 542 ACACTATAGGACCTTAATGGGTGGCGGGTGGGAGACCGACTCATTTGATTCTTG 483
QY 568 AAGGCTTCCTGGGCTCAGGACAGTAAATGCCCCGCTGGATCTGATGCTTCCGG 627
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 482 AAGGCTTCCTGGGCTCAGGACAGTAAATGCCCCGCTGGATCTGATGCTTCCGG 423
QY 628 ATTGGTGGCCAGTCCGCGGGGTGTAGATGTTCTCTGACGGCCCCAAGGGTGCCTGAAC 687
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 422 ATTGGTGGCCAGTCCGCGGGGTGTAGATGTTCTCTGACGGCCCCAAGGGTGCCTGAAC 363
QY 688 CCGCGGTGACCTCTCTCAGGAGACCTTCCAGTGCAGACCTTCTTCTATGATGACG 747
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 362 CCGCGGTGACCTCTCTCAGGAGACCTTCCAGTGCAGACCTTCTTCTATGATGACG 303
QY 748 ACGGGGGCCCCGGTGAAGAGGGGTCCCGTGGGTGATACAAACAGACGCTTTCAGCAGC 807
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 302 ACGGGGGCCCCGGTGAAGAGGGGTCCCGTGGGTGATACAAACAGACGCTTTCAGCAGC 243
QY 808 GGCTGAGACAGGTGGTGGACCTGGCGCTGCTGCCCTCATTTCCCGCTGGCGCGCGC 867
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 242 GGCTGAGACAGGTGGTGGACCTGGCGCTGCTGCCCTCATTTCCCGCTGGCGCGCGC 183
QY 868 CTGAGCTCGCTGCTTGGCGGTGGAGGACCTCCGCTGTGCCAGCGGCTTACCGCAGCC 927
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 CTGAGCTCGCTGCTTGGCGGTGGAGGACCTCCGCTGTGCCAGCGGCTTACCGCAGCC 123
QY 928 AGGCGCGGGGATCGCTCTCTG 948
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 AGGCGCGGGGATCGCTCTCTG 102

```

RESULT 5  
BM833169/c 486 bp mRNA linear EST 06-MAR-2002  
LOCUS K-ESP0107772 SSSNU484s1 Homo sapiens cDNA clone SSSNU484s1-12-C04  
DEFINITION 5', mRNA sequence.  
ACCESSION BM833169  
VERSION BM833169.1 GI:19189578  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 486)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.

```

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: Yongsung@mail.kr.ibm.re.kr
Plate: 12 row: C column: 04
High quality sequence stop: 486.
FEATURES
source
location/Qualifiers
1..486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="SSSNU484s1-12-C04"
/clone_lib="SSSNU484s1"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/cell_line="SNU-484"
/lab_host="Top10F"
/notes="Organ: Stomach; Vector: pTZ18Rp1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was decapped with tobacco
acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
including EcoRI site by treatment of T4 RNA ligase. The
first strand cDNA was synthesized from oligo dT-selected
mRNA by priming with dr-tailed vector. The dr-tailed
vector was adjusted to have about 60nt. The cDNA vector
was circularized with E. coli DNA ligase after digestion
of EcoRI which site is also included in vector. An RNA
strand converted to a DNA strand by Okayama-Berg method.
The obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
After analyzing and sequencing about 2,000 - 3,000
colonies in original cDNA library, the abundant cDNAs were
selected and amplified by PCR reaction using vector region
primer including T7 promoter as 5' primer and N(orf)14 as
3' primer. The PCR products were used as template for
synthesis of biotinylated single stranded RNA by in vitro
transcription reaction. The synthesized RNA probes were
hybridized with antisense single stranded cDNAs prepared
from original library and incubated with avidin-gel.
After removing DNA-RNA hybrids by centrifuge, the
subtracted cDNA libraries were constructed by
transformation of the remaining DNA into competent cells E.
coli Top10F' with electroporation method."
BASE COUNT      141 a      113 c      132 g      100 t
ORIGIN
Query Match      34.0%; Score 470; DB 14; Length 486;
Best Local Similarity 100.0%; Pred. No. 3,1e-229;
Matches 470; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 264 TTGTGTCACCTTACCTTTGCAAGGATACCTTTTATTTCTTTAAGATCTCTGTGTTAT 323
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 470 TTGTGTCACCTTACCTTTGCAAGGATACCTTTTATTTCTTTAAGATCTCTGTGTTAT 411
QY 324 ACACAGATTTTAACTTATGCTCTGCTGACCACTGAAATCTCTTCCAGTCACAG 383
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 410 ACACAGATTTTAACTTATGCTCTGCTGACCACTGAAATCTCTTCCAGTCACAG 351
QY 384 TGTCAACCTTACCCCACTGCAAGAGAGTTTGTAGGGGCAATCAACACCGAGAA 443
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 350 TGTCAACCTTACCCCACTGCAAGAGAGTTTGTAGGGGCAATCAACACCGAGAA 291
QY 444 GTCAAGCCCTCAACCACTGAGGTGTGGGGGTAGGATCTGCATTTCTTATATCAA 503
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 290 GTCAAGCCCTCAACCACTGAGGTGTGGGGGTAGGATCTGCATTTCTTATATCAA 231
QY 504 CCCCACTATATAGGGACCTAAATGGGTGGGGGTGGGGAGACCGACTCATTTGATTT 563
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 230 CCCCACTATATAGGGACCTAAATGGGTGGGGGTGGGGAGACCGACTCATTTGATTT 171

```

```

OY 564 CTTGAAGGCTTCTGGCTTCACGACGTAATATGCCCCGCTGCGATCTGTAGCTT 623
    |||||||
Db 170 CTTGAAGGCTTCTGGCTTCACGACGTAATATGCCCCGCTGCGATCTGTAGCTT 111
OY 624 CCGGATTCGGTGGCCAGTCCGCGGGGTGTAGATGTTCTTCAGCGCCCCCAAGGGTGCCTG 683
    |||||||
Db 110 CCGGATTCGGTGGCCAGTCCGCGGGGTGTAGATGTTCTTCAGCGCCCCCAAGGGTGCCTG 51
OY 684 AACGCGCGGCTACCTCTCTCAGGAAGACTTGAAGCTGGACACTTCT 733
    |||||||
Db 50 AACGCGCGGCTACCTCTCTCAGGAAGACTTGAAGCTGGACACTTCT 1

RESULT 6
BO650059/c 994 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT_8297859 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6269873
DEFINITION 5', mRNA sequence.
ACCESSION BO650059
VERSION BO650059.1 GI:21774231
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 994)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2444 row: b column: 18
High quality sequence stop: 487.
Location/Qualifiers
    1..994
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_1fb="NIH_MGC_100"
    /clone_1lb="NIH_MGC_100"
    /issue_type="hepatocellular carcinoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
    EcoRI; CDNA made by oligo-dT priming. Directionally cloned
    into EcoRI/XhoI sites using the following 5' adaptor:
    GGCACGAG(G). Size-selected >500bp for average insert size
    1.8kb. Library constructed by Ling Hong in the laboratory
    of Gerald M. Rubin (University of California, Berkeley)
    using ZAP-cDNA synthesis kit (Stratagene) and Superscript
    II RT (Life Technologies). Note: this is a NIH_MGC
    library."
BASE COUNT 137 a 423 c 234 g 200 t
ORIGIN

Query Match 30.0%; Score 414; DB 14; Length 994;
Best Local Similarity 100.0%; Pred. No. 1.5e-200; Indels 0; Gaps 0;
Matches 414; Conservative 0; Mismatches 0;

OY 535 CCGTGGGGAGACGACTACTGAGTTCTTGAAGCCTTCGGGCTCCAGCAGCTAA 594
    |||||||
Db 514 CCGTGGGGAGACGACTACTGAGTTCTTGAAGCCTTCGGGCTCCAGCAGCTAA 455
OY 595 TTGCCCCCGCTGTGATCTGCTAGCTTCGGAATTCGGTGGCCAGTCCGGGGGTGTAG 654
    |||||||
Db 454 TTGCCCCCGCTGTGATCTGCTAGCTTCGGAATTCGGTGGCCAGTCCGGGGGTGTAG 395
```

```

OY 655 ATGTTCTGACGGGCCCCAAAGGGTGCCTGAACCGCGCGGCTACACTCTTCAGGAAGACT 714
    |||||||
Db 394 ATGTTCTGACGGGCCCCAAAGGGTGCCTGAACCGCGCGGCTACACTCTTCAGGAAGACT 335
OY 715 TCGAAGCTGACACCTTCTTCATGATGATGACAGCGCGCGCCCGGTGAAGGGTCC 774
    |||||||
Db 334 TCGAAGCTGACACCTTCTTCATGATGATGACAGCGCGCGCCCGGTGAAGGGTCC 275
OY 775 CCGTGGGTTACACAGCAGCTTCACAGCGGGGTGAACAGAGTGGCGACCTGGCG 834
    |||||||
Db 274 CCGTGGGTTACACAGCAGCTTCACAGCGGGGTGAACAGAGTGGCGACCTGGCG 215
OY 835 CTGCTGCGGCTCATCTTCCCGCTGGCGCGGCTCAGCTGCTGCTTCGCTGGAGG 894
    |||||||
Db 214 CTGCTGCGGCTCATCTTCCCGCTGGCGCGGCTCAGCTGCTGCTTCGCTGGAGG 155
OY 895 CACCTCGCTGTCCACAGCGGCTACCGCAGCGGCGGGGATCGCTCTG 948
    |||||||
Db 154 CACCTCGCTGTCCACAGCGGCTACCGCAGCGGCGGGGATCGCTCTG 101

RESULT 7
BM923267/c 984 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT_6626059 NIH_MGC_116 Homo sapiens CDNA clone IMAGE:5758845
DEFINITION 5', mRNA sequence.
ACCESSION BM923267
VERSION BM923267.1 GI:19373646
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 984)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12803 row: e column: 22
High quality sequence stop: 673.
Location/Qualifiers
    1..984
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_1fb="IMAGE:5758845"
    /clone_1lb="NIH_MGC_116"
    /lab_host="DH10B"
    /note="Organ: pooled colon, kidney, stomach; Vector:
    pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
    source anonymous pool of 3 colons, age 26 yo male, 49 yo
    female, 71 yo male colon; 46 yo male kidney, and pool of 2
    stomachs, 62 yo male and 70 yo female. Library is
    oligo-dT primed and directionally cloned (EcoRV site is
    destroyed upon cloning). Average insert size 1.4 kb,
    insert size range 1-3 kb. Library is normalized and
    enriched for full-length clones and was constructed by C.
    Gruber (Invitrogen). Research Genetics tracking code
    023. Note: this is a NIH_MGC library."
BASE COUNT 246 a 254 c 279 g 205 t
ORIGIN

Query Match 24.7%; Score 342; DB 14; Length 984;
Best Local Similarity 99.7%; Pred. No. 1.2e-163; Indels 0; Gaps 0;
Matches 392; Conservative 0; Mismatches 1;
```

QY 556 TTGAGTTCTTGAAGCCTCTCTGCGCTCCAGCCAGTATTGCCGCCGCTCTGATCTGG 615  
|||||  
Db 515 TTGAGTTCTTGAAGGCTTCTCTGCGCTCCAGCCAGTATTGCCGCCGCTCTGATCTGG 456  
QY 616 TCTAGCTTCCGATTCGGTGGCAGTCCGCGGGGGTGTAGATGTTCTTACGAGCCCAAG 675  
|||||  
Db 455 TCTAGCTTCCGATTCGGTGGTCACTCCGCGGGGGTGTAGATGTTCTTACGAGCCCAAG 396  
QY 676 GGTGCTGAAGCGCCGCGGTCACTCTCTTCAAGAACTTGAAGCTGGACACCTTCTTC 735  
|||||  
Db 395 GGTGCTGAAGCGCCGCGGTCACTCTCTTCAAGAACTTGAAGCTGGACACCTTCTTC 336  
QY 736 TCAATGATGAGACGCGCGCGCGCTAGAAAGGGTCTCCGCTTCCGCTACACAGCAG 795  
|||||  
Db 335 TCAATGATGAGACGCGCGCGCGCTAGAAAGGGTCTCCGCTTCCGCTACACAGCAG 276  
QY 796 CTCTTCAAGACGGGCTGAGACAGGTGAGTGGCTGAGCTGGCGCTGCTCATCTTCCCG 855  
|||||  
Db 275 CTCTTCAAGACGGGCTGAGACAGGTGAGTGGCTGAGCTGGCGCTGCTCATCTTCCCG 216  
QY 856 GCTGCGCGCGCTGAGTGGCTGAGTGGCTGAGCTGGCGCTGCTCATCTTCCCG 915  
|||||  
Db 215 GCTGCGCGCGCTGAGTGGCTGAGTGGCTGAGCTGGCGCTGCTCATCTTCCCG 156  
QY 916 CTCACCGCACCAGGGCGCGGGATGCGCTCTG 948  
|||||  
Db 155 CTCACCGCACCAGGGCGCGGGATGCGCTCTG 123

RESULT 8  
LOCUS BG749180 1045 bp mRNA linear EST 15-MAY-2001  
DEFINITION 602708139F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:484474 5',  
mRNA sequence.  
ACCESSION BG749180  
VERSION BG749180.1 GI:14059833  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1045)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1682 row: n column: 09  
High quality sequence stop: 822.

## FEATURES

source

1. 1045  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:484474"  
/clone\_id="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pORF7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCAGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC Library. !"

BASE COUNT

245 a 322 c 300 g 178 t

## ORIGIN

Query Match 22.3%; Score 308; DB 12; Length 1045;  
Best Local Similarity 99.6%; Pred. No. 3.1e-146;  
Matches 548; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 522 CTAAATGGGTGGGGGGGGGAGACGACTTGTGATGTTCTTGAAGGCTTCTGACC 581  
|||||  
Db 2 CTAATGGGTGGGGGGGGGAGACGACTTGTGATGTTCTTGAAGGCTTCTGACC 61  
QY 582 TCCAGCCAGTAAATGGCCCGGCTGTGGATGTGGTCACTTCCGAAATGCGGGCCAGT 641  
|||||  
Db 62 TCCAGCCAGTAAATGGCCCGGCTGTGGATGTGGTCACTTCCGAAATGCGGGCCAGT 121  
QY 642 CCGGGGGGTGTAGATGTTCTTCAAGCGCCCAAGAGGTGCTGAAAGCGCGCGCTCACCTC 701  
|||||  
Db 122 CCGGGGGGTGTAGATGTTCTTCAAGCGCCCAAGAGGTGCTGAAAGCGCGCGCTCACCTC 181  
QY 702 CTTAGGAAAGCTTCAAGCTGAGACCTTCTTCTCATGATGACAGCGCGCGCCCGC 761  
|||||  
Db 182 CTTAGGAAAGCTTCAAGCTGAGACCTTCTTCTCATGATGACAGCGCGCGCCCGC 241  
QY 762 GTAGAAAGGGTCCCGGCTGCGTACACAGCAGCTTTCACGAGCGGCTGAGACAGT 821  
|||||  
Db 242 GTAGAAAGGGTCCCGGCTGCGTACACAGCAGCTTTCACGAGCGGCTGAGACAGT 300  
QY 822 GCTGAGACCTGGCGCTGCTGCGCTCATCTTCCCGCTGGCGCGCGCTCAGCTGCTGCT 881  
|||||  
Db 301 GCTGAGACCTGGCGCTGCTGCGCTCATCTTCCCGCTGGCGCGCGCTCAGCTGCTGCT 360  
QY 882 TCGGCTGGGAGGACCTTCCGCTGAGCGGCTTCCAGCGGCTTCCAGCGGCTTCCAGCGG 941  
|||||  
Db 361 TCGGCTGGGAGGACCTTCCGCTGAGCGGCTTCCAGCGGCTTCCAGCGGCTTCCAGCGG 419  
QY 942 CCTCCTGAAGAGAGAGAACTGACGATCTCAGAGGTGAAGAGAAAGTAAAGGCGCTGAC 1001  
|||||  
Db 420 CCTCCTGAAGAGAGAGAACTGACGATCTCAGAGGTGAAGAGAAAGTAAAGGCGCTGAC 479  
QY 1002 GCTAGAGCGTCCACCCAGAGAGACACTAGAGCTTGCAGGACTTGCAGTACGCTCAA 1061  
|||||  
Db 480 GCTAGAGCGTCCACCCAGAGAGACACTAGAGCTTGCAGGACTTGCAGTACGCTCAA 539  
QY 1062 GTTTTTCACC 1071  
|||||  
Db 540 GTTTTTCACC 549

## RESULT 9

B0644671/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 929)  
Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: CCAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM2497 row: j column: 18



High quality sequence stop: 582.

Location/Qualifiers

FEATURES

SOURCE

1. 929

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6293873"

/clone.lib="NIH\_MGC\_100"

/tissue.type="hepatocellular carcinoma, cell line"

/lab.host="DH10B (phage-resistant)"

/note="Organ: Liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGAG(G). Size-selected &gt;500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

BASE COUNT 283 a 212 c 238 g 196 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 5.5e-136;

Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 929;

100.0%; Pred. No. 5.5e-136;

Location/Qualifiers

FEATURES

SOURCE

1. 939

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6284804"

/clone.lib="NIH\_MGC\_100"

/tissue.type="hepatocellular carcinoma, cell line"

/lab.host="DH10B (phage-resistant)"

/note="Organ: Liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGAG(G). Size-selected &gt;500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

BASE COUNT 283 a 212 c 240 g 202 t 2 others

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 5.5e-136;

Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

20.8%; Score 288; DB 14; Length 939;

100.0%; Pred. No. 5.5e-136;

Location/Qualifiers

FEATURES

SOURCE

1. 939

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6284804"

/clone.lib="NIH\_MGC\_100"

/tissue.type="hepatocellular carcinoma, cell line"

/lab.host="DH10B (phage-resistant)"

/note="Organ: Liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGAG(G). Size-selected &gt;500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."&lt;/



FEATURES  
source

Location/Qualifiers  
1.357  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:272737"  
/clone.lib="NCI\_CGAP\_Sub4"  
/lab.host="DHIOB (Life Technologies)"  
/note="vector: p7773D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not 1; Site\_2: Eco RI; The  
NCI\_CGAP\_Sub4 library is a subtracted library derived from  
the NCI\_CGAP\_Sub2 library which is a subtracted library  
derived from the NCI\_CGAP\_Sub1 library, which is a  
subtracted library derived from B1. B1 constitutes a  
mixture of 21 normalized or subtracted NCI\_CGAP  
libraries: NCI\_CGAP\_Co4, NCI\_CGAP\_Pr22, NCI\_CGAP\_Pr28,  
NCI\_CGAP\_Co10, NCI\_CGAP\_Co16, NCI\_CGAP\_Kid5,  
NCI\_CGAP\_Kid12, NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid11,  
NCI\_CGAP\_Lym2, NCI\_CGAP\_Pr2, NCI\_CGAP\_Co8, NCI\_CGAP\_CLL1,  
NCI\_CGAP\_Lu12, NCI\_CGAP\_Brn23, NCI\_CGAP\_Lu5,  
NCI\_CGAP\_Lu24, NCI\_CGAP\_Lu19, NCI\_CGAP\_Gc4, NCI\_CGAP\_Gc6,  
NCI\_CGAP\_Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a driver whose composition is detailed below:  
NCI\_CGAP\_Kid3 pool 1 : LHAM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonoids 1322376-1323911,  
1456008-1456775, 1500552-1502855) NCI\_CGAP\_Kid5 pool 1 :  
LHAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids  
1323912-1325831, 1471368-1472903, 1492104-1493255)  
NCI\_CGAP\_Lu5 pool 1 : LHAM 3575-3582, 3851-3854 (IMAGE  
Clonoids 1414920-1417991, 1520904-1522439) NCI\_CGAP\_Gc4  
pool 1 : LHAM 3164-3167, 3716-3720, 3733-3735 (IMAGE  
Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743  
) NCI\_CGAP\_Pr22 pool 1 : LHAM 2457-2459, 2758-2759,  
3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959,  
1217928-1220615) NCI\_CGAP\_Co10 pool 1 : LHAM 2644-2653,  
2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351  
) Subtraction was performed as previously described  
(Bonaldo, Lennon & Soares (1996): Normalization and  
Subtraction: Two Approaches to Facilitate Gene Discovery.  
Genome Research 6, 791-806.)  
TAG\_Lib=NCI\_CGAP\_Kid3  
TAG\_Tissue=Kidney  
TAG\_SEO=ATGTC"

BASE COUNT 73 a 92 c 77 g 115 t  
ORIGIN

Query Match 14.2%: Score 196; DB 10; Length 357;  
Best Local Similarity 100.0%: Pred. No. 7.2e-89;  
Matches 196: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 TTGAGTTTGTGAGGCTTCCTGCGCTCCAGCCAGCTAAATGGCCCCGCTGTGAGATCTGG 615  
|||||  
DB 162 TTGAGTTTCTTAAGGCTTCCTGCGCTCCAGCCAGCTAAATGGCCCCGCTGTGATCTGG 221  
QY 616 TCTAGCTTCGGATTGCGGCGCAGTCCGCGGCTGTAAATGTTCTGTGACGGCCCCAAG 675  
|||||  
DB 222 TCTAGCTTCGGATTGCGGCGCAGTCCGCGGCTGTAAATGTTCTGTGACGGCCCCAAG 281  
QY 676 GGTGCTTGAACGCCGCGGTCACTCTTCAGAGAACTTCGAGAGCTTGACACTTCTTC 735  
|||||  
DB 282 GGTGCTTGAACGCCGCGGTCACTCTTCAGAGAACTTCGAGAGCTTGACACTTCTTC 341  
QY 736 TCATGTATGACGACGC 751  
|||||  
DB 342 TCATGTATGACGACGC 357  
|||||  
RESULT 14 364 bp mRNA linear EST 24-NOV-2000  
LOCUS BF364659  
DEFINITION MR2-NN1111-070800-007-h12 NN1111 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF364659

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

BF364659.1 GI:11326684  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 364)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Birones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&lc2=MR2-NN1111-  
070800-007-h12&lc3=2000-08-07&lc4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 9.

FEATURES  
source

Location/Qualifiers  
1.364  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone.lib="NN1111"  
/dev.stage="Adult"  
/note="organ: nervous normal. Vector: puc18. Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 82 a 114 c 112 g 56 t  
ORIGIN

Query Match 10.3%: Score 142; DB 12; Length 364;  
Best Local Similarity 100.0%: Pred. No. 3.4e-61;  
Matches 142: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 930 GCGCGGAGATGCGCTCTCTGAAGACGAGAACTGACGAAATCCACAGGTGAAGAAG 989  
|||||  
DB 87 GCGCGGAGATGCGCTCTCTGAAGACGAGAACTGACGAAATCCACAGGTGAAGAAG 146  
QY 990 TAACGGCCGTGCGCTTACGCGCTCCAGCAGAGAGACACTAGAGAGCTTGACAGACTCGGA 1049  
|||||  
DB 147 TAACGGCCGTGCGCTTACGCGCTCCAGCAGAGAGACACTAGAGAGCTTGACAGACTCGGA 206  
QY 1050 GTAGACGCTCAAGTTTTCAC 1071  
|||||  
DB 207 GTAGACGCTCAAGTTTTCAC 228  
|||||  
RESULT 15 835 bp mRNA linear EST 15-MAY-2001  
LOCUS BG750306/c  
DEFINITION 602709116P1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4845915 5',  
mRNA sequence.  
ACCESSION BG750306  
VERSION BG750306.1 GI:14060959  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 835)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D.

## FEATURES

## Source

BASE COUNT	216 a	169 c	245 g	205 t
ORIGIN				

Query Match	10.3%	Score 142;	DB 12;	Length 835;
Best Local Similarity	99.5%;	Pred. No. 3.5e-61;		
Matches 192;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0

QY	753	GCCTCCGCGCGTAAAGAGGGGGTCCCGTTGGCGGTAAACAAGACAGCGTCTTCAAGAGAGGGCTG	81.2
Db	813	GGGGCCCGCGTAAAGAGGGGGTCCCGCTTGGCGGTAAACAAGACAGCGTCTTCAAGAGAGGGCTG	75.4
QY	813	AGACAGAGTGGCTGAGACCTGGCGGCTGCTGCCTGCATCTTCCCGCGCTGGCGCGCGCTCAG	87.7
Db	753	AGACAGAGTGGTGGAGACCTGGCGGCTGCTGCCTGCATCTTCCCGCGCTGGCGCGCGCTCAG	69.4
QY	873	CTCGCTGCTTGGCGCTCGGGGAGGACACTCCGCTGTGCCAGCGGGCTCACCGACCCAGGGC	93.2
Db	693	CTCGCTGCTTGGCGCTCGGGGAGGACACTCCGCTGTGCCAGCGGGCTCACCGACCCAGGGC	63.4
QY	933	GGGGATCGCCCTC	94.5
Db	633	GGGGATCGCCCTC	62.1

Search completed: January 18, 2003, 23:34:29  
Job time : 2221 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 21:56:44 ; Search time 119 Seconds  
(without alignments)  
5178.067 Million cell updates/sec

Title: US-09-674-593-1  
Perfect score: 1382  
Sequence: 1 cattatgctaacagcataaa.....taaagcaattaagtctctgg 1382

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 393868 seqs, 222934149 residues

Word size : 0 787736

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	2.0	2116	US-09-925-297-261	Sequence 261, App
c 2	27	2.0	397658	US-09-813-320-3	Sequence 3, Appli
3	25	1.8	447	US-09-954-531-742	Sequence 742, App
c 4	25	1.8	511	US-09-954-456-222	Sequence 222, App
c 5	25	1.8	511	US-09-954-456-394	Sequence 394, App
c 6	25	1.8	38844	US-10-060-333-3	Sequence 3, Appli
c 7	25	1.8	143306	US-09-729-920-3	Sequence 3, Appli
c 8	24	1.7	359	US-09-969-373-1071	Sequence 1071, Ap
9	24	1.7	401	US-09-946-807-447	Sequence 447, App
10	24	1.7	401	US-09-946-807-448	Sequence 448, App
11	24	1.7	401	US-09-795-668-447	Sequence 447, App
12	24	1.7	401	US-09-795-668-448	Sequence 448, App
13	24	1.7	401	US-09-795-686-447	Sequence 447, App
14	24	1.7	401	US-09-795-686-448	Sequence 448, App
15	24	1.7	433	US-09-960-352-12891	Sequence 12891, A
c 16	24	1.7	501	US-09-954-531-726	Sequence 726, App
17	24	1.7	933	US-09-817-607-55	Sequence 55, Appl
18	24	1.7	11270	US-09-977-221-3	Sequence 3, Appli
19	24	1.7	45845	US-09-927-091-6	Sequence 6, Appli

c 20	24	1.7	49744	10	US-09-927-091-4	Sequence 4, Appli
21	24	1.7	56737	10	US-09-782-378A-17	Sequence 17, Appl
22	24	1.7	173808	12	US-10-003-806-10	Sequence 10, Appl
23	24	1.7	249487	9	US-10-026-188-3	Sequence 3, Appli
24	24	1.7	368004	10	US-09-949-654-3	Sequence 3, Appli
c 25	24	1.7	368004	10	US-09-949-654-3	Sequence 3, Appli
c 26	24	1.7	659158	9	US-09-771-208-20	Sequence 20, Appl
27	24	1.7	1503841	9	US-09-946-807-1	Sequence 1, Appli
28	24	1.7	1503841	10	US-09-795-668-1	Sequence 1, Appli
29	24	1.7	1503841	10	US-09-795-668-1	Sequence 1, Appli
c 30	23	1.7	73	10	US-09-920-300A-1268	Sequence 1268, Ap
c 31	23	1.7	73	12	US-10-033-528-1268	Sequence 1268, Ap
c 32	23	1.7	231	9	US-09-920-455-94	Sequence 94, Appl
c 33	23	1.7	288	9	US-10-025-380-284	Sequence 284, App
c 34	23	1.7	288	9	US-10-025-380-751	Sequence 751, App
c 35	23	1.7	288	10	US-09-922-217-284	Sequence 284, App
c 36	23	1.7	288	10	US-09-922-217-751	Sequence 751, App
c 37	23	1.7	288	10	US-09-833-263-284	Sequence 284, App
c 38	23	1.7	288	10	US-09-833-263-751	Sequence 751, App
c 39	23	1.7	299	9	US-10-040-739-369	Sequence 369, App
c 40	23	1.7	399	10	US-09-764-869-2028	Sequence 2028, Ap
41	23	1.7	399	10	US-09-764-869-2029	Sequence 2029, Ap
c 42	23	1.7	422	10	US-09-960-352-6063	Sequence 6063, Ap
c 43	23	1.7	429	9	US-10-040-739-197	Sequence 197, App
c 44	23	1.7	476	10	US-09-917-800A-382	Sequence 382, App
c 45	23	1.7	478	10	US-09-954-456-1350	Sequence 1350, Ap

ALIGNMENTS

RESULT 1  
US-09-925-297-261  
; Sequence 261, Application US/09925297  
; Patent No. US20020081659A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PAL05  
; CURRENT APPLICATION NUMBER: US/09/925,297  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05989  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 928  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 261  
; LENGTH: 2116  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (7)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (16)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (25)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (35)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-297-261

Query Match 2.0%; Score 28; DB 10; Length 2116;  
Best Local Similarity 100.0%; Pred. No. 0.0003;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1225 CGCACACACACACACACAAATATG 1252  
|||||  
Db 204 CGCACACACACACACACAAATATG 231

RESULT 2  
US-09-813-320-3/c  
; Sequence 3, Application US/09813320  
; Patent No. US20020142378A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHANG, Hongyu et al.  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; FILE REFERENCE: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CL001172  
; CURRENT APPLICATION NUMBER: US/09/813,320  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 397658  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(397658)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-813-320-3

Query Match 2.0%; Score 27; DB 10; Length 397658;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1224 ACCACACACACACACACAAATA 1250  
|||||  
Db 297968 ACCACACACACACACACAAATA 297942

RESULT 3  
US-09-954-531-742  
; Sequence 742, Application US/09954531  
; Patent No. US20020165180A1  
; GENERAL INFORMATION:  
; APPLICANT: Weaver, Zoe  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
; FILE REFERENCE: Gene Sets  
; FILE REFERENCE: 689290-77  
; CURRENT APPLICATION NUMBER: US/09/954,531  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US/60/233,133  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,009  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,034  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,509  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US/60/234,567  
; PRIOR FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 1392  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 742  
; LENGTH: 447  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-531-742

Query Match 1.8%; Score 25; DB 9; Length 447;  
Best Local Similarity 100.0%; Pred. No. 0.0091;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAATAT 1251  
|||||

Db 395 CACACACACACACACAAATAT 419

RESULT 4  
US-09-954-456-222/c  
; Sequence 222, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; FILE REFERENCE: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 222  
; LENGTH: 511  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-222

Query Match 1.8%; Score 25; DB 10; Length 511;  
Best Local Similarity 100.0%; Pred. No. 0.0092;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAATAT 1251  
|||||  
Db 402 CACACACACACACACAAATAT 378

RESULT 5  
US-09-954-456-394/c  
; Sequence 394, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; FILE REFERENCE: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25

; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 394  
; LENGTH: 511  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: n-a,t,g or c  
US-09-954-456-394

Query Match 1.8%; Score 25; DB 10; Length 511;  
Best Local Similarity 100.0%; Pred. No. 0.0092;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1227 CACACACACACACACAATAT 1251  
|||||  
Db 402 CACACACACACACACAATAT 378

RESULT 6  
US-10-060-333-3/c  
; Sequence 3, Application US/10060333  
; Patent No. US20020115186A1  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO00862DIV  
; CURRENT APPLICATION NUMBER: US/10/060,333  
; CURRENT FILING DATE: 2002-08-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 38844  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-060-333-3

Query Match 1.8%; Score 25; DB 12; Length 38844;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1227 CACACACACACACACAATAT 1251  
|||||  
Db 16438 CACACACACACACACAATAT 16414

RESULT 7  
US-09-729-920-3/c  
; Sequence 3, Application US/09729920  
; Patent No. US20020103115A1  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLER, Karl et al  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CLO00858  
; CURRENT APPLICATION NUMBER: US/09/729,920  
; CURRENT FILING DATE: 2000-12-06

; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 143306  
; TYPE: DNA  
; ORGANISM: Human  
US-09-729-920-3

Query Match 1.8%; Score 25; DB 10; Length 143306;  
Best Local Similarity 100.0%; Pred. No. 0.013;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1227 CACACACACACACACAATAT 1251  
|||||  
Db 18457 CACACACACACACACAATAT 18433

RESULT 8  
US-09-969-373-1071/c  
; Sequence 1071, Application US/09969373  
; Patent No. US20020133852A1  
; GENERAL INFORMATION:  
; APPLICANT: Effertz, Roger J.  
; APPLICANT: Hauge, Brian M.  
; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping  
; FILE REFERENCE: 38-10(52679)A  
; CURRENT APPLICATION NUMBER: US/09/969,373  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US 09/754,853  
; PRIOR FILING DATE: 2001-01-05  
; PRIOR APPLICATION NUMBER: US 09/760,427  
; PRIOR FILING DATE: 2001-01-13  
; PRIOR APPLICATION NUMBER: US 09/855,768  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 4593  
; SEQ ID NO 1071  
; LENGTH: 359  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-09-969-373-1071

Query Match 1.7%; Score 24; DB 10; Length 359;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1227 CACACACACACACACAATA 1250  
|||||  
Db 65 CACACACACACACACAATA 42

RESULT 9  
US-09-946-807-447  
; Sequence 447, Application US/09946807  
; Patent No. US20020165144A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinhorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/946,807  
; CURRENT FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: US/09/795,668  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 447  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-946-807-447

Query Match 1.7%; Score 24; DB 9; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAATA 1250  
|||||  
DB 263 CACACACACACACACACAATA 286

## RESULT 10

US-09-946-807-448  
; Sequence 448, Application US/09946807  
; Patent No. US20020165144A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/946,807  
; PRIOR FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: US/09/795,668  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 448  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-946-807-448

Query Match 1.7%; Score 24; DB 9; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAATA 1250  
|||||  
DB 262 CACACACACACACACACAATA 285

## RESULT 11

US-09-795-668-447  
; Sequence 447, Application US/09795668  
; Patent No. US20020045577A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/795,668  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 447  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-795-668-447

Query Match 1.7%; Score 24; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAATA 1250  
|||||  
DB 263 CACACACACACACACACAATA 286

## RESULT 12

US-09-795-668-448  
; Sequence 448, Application US/09795668  
; Patent No. US20020045577A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/795,668  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 448  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-795-668-448

Query Match 1.7%; Score 24; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAATA 1250  
|||||  
DB 262 CACACACACACACACACAATA 285

## RESULT 13

US-09-795-686-447  
; Sequence 447, Application US/09795686  
; Patent No. US20020094954A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2005-001  
; CURRENT APPLICATION NUMBER: US/09/795,686  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,715  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 447  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-795-686-447

Query Match 1.7%; Score 24; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAATA 1250  
|||||  
DB 263 CACACACACACACACACAATA 286

## RESULT 14

US-09-795-686-448  
; Sequence 448, Application US/09795686  
; Patent No. US20020094954A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2005-001



; CURRENT APPLICATION NUMBER: US/09/795,686  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,715  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 448  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-795-686-448

Query Match 1.7%; Score 24; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAATA 1250  
|||||  
DB 262 CACACACACACACACAAATA 285

RESULT 15  
US-09-960-352-12891  
; Sequence 12891, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 12891  
; LENGTH: 433  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 55-LIB3058-019-Q1-K1-F12  
US-09-960-352-12891

Query Match 1.7%; Score 24; DB 10; Length 433;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACAAATA 1250  
|||||  
DB 110 CACACACACACACACAAATA 133

Search completed: January 18, 2003, 23:44:45  
Job time : 521 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 19:22:24 ; Search time 265 seconds  
(without alignments)

Title: US-09-674-593-1  
Perfect score: 1382  
Sequence: 1 cattatgctaacagcataaa.....taagcaattaagtctctgg 1382  
11744.384 Million cell updates/sec

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1382	100.0	1382	21 AAZ36643	Human tumour rejec
2	595	43.1	2167	21 AAZ36644	Human tumour rejec
3	498	36.0	1956	22 AAI58192	Human polynucleoti
4	241	17.4	246	21 AAZ36649	EST AAG63443 which
5	139	10.1	492	22 AAI59978	Human polynucleoti
6	60	4.3	60	24 ABN38745	Human spliced tran
7	28	2.0	2116	21 AAC99033	Human pancreatic c
8	28	2.0	2116	22 AAS00847	Human cDNA clone H
9	28	2.0	2219	22 ABA07127	Human pancreatic c

10	28	2.0	2219	22 AAK89341	Human digestive sy
11	28	2.0	165199	24 ABK83460	Human cDNA differe
12	26	1.9	383	22 AAI82036	Human polynucleoti
13	26	1.9	1747	22 AAD05327	Human secreted pro
14	26	1.9	16584	22 AAK83864	Human immune/haema
15	25	1.8	25	21 AAZ36647	PCR primer VDE119
16	25	1.8	25	21 AAZ36648	PCR primer VDE120
17	25	1.8	186	21 ABN81170	Shrimp polynucleot
18	25	1.8	393	21 AAA74261	Lobolily pine SSR
19	25	1.8	407	22 AAI90643	Human polynucleoti
20	25	1.8	447	24 ABL63472	Breast cancer rela
21	25	1.8	498	24 ABQ57773	Human colon cancer
22	25	1.8	499	22 AAI86291	Human polynucleoti
23	25	1.8	511	24 ABL64912	Lung cancer relate
24	25	1.8	511	24 ABL65084	Lung cancer relate
25	25	1.8	577	21 AAA74262	Lobolily pine SSR
26	25	1.8	612	23 ABV59319	Human prostate exp
27	25	1.8	860	24 ABL88516	C geographus mu-co
28	25	1.8	1347	22 AAI93733	Human polynucleoti
29	25	1.8	1559	24 ABK83492	Human cDNA differe
30	25	1.8	2215	23 AAS82731	DNA encoding novel
31	25	1.8	2319	19 AAV64070	Arabidopsis thalia
32	25	1.8	2478	22 AAI7764	Human cDNA sequenc
33	25	1.8	5371	24 ABL34295	Human immune syste
34	25	1.8	5919	22 AAS46349	Tumour suppressor
35	25	1.8	5919	24 ABL70213	Chemically treated
36	25	1.8	5919	24 ABL32732	Human immune syste
37	25	1.8	5919	24 AAS61168	Human gene regulat
38	25	1.8	5919	24 ABK31258	Signal transductio
39	25	1.8	5936	24 ABL33651	Human immune syste
40	25	1.8	6673	24 ABQ67112	Human angiogenesis
41	25	1.8	8201	21 ABA88864	Human dentin sialo
42	25	1.8	8201	24 ABQ73537	Human dentin sialo
43	25	1.8	10552	24 ABL70387	Chemically treated
44	25	1.8	10552	24 AAS61339	Human gene regulat
45	25	1.8	10552	24 ABK31426	Signal transductio

ALIGNMENTS

RESULT 1  
AAZ36643  
ID AAZ36643 standard; cDNA; 1382 BP.  
XX  
AC AAZ36643;  
XX  
DT 22-FEB-2000 (first entry)  
XX Human tumour rejection antigen RUR-1 antisense cDNA sequence.  
DE Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT CDS 738..992  
FT FT /\*tag= a  
FT FT 523..547  
FT FT /\*tag= b  
FT FT /note= "binding site for primer VDE119 (see AAZ36647)"  
FT FT /\*tag= c  
FT FT /note= "binding site for primer VDE120 (see AAZ36648)"  
XX WO9958546-A1.  
XX 18-NOV-1999.  
XX 13-MAY-1999;  
XX 99WO-US10424.

PR	13-MAY-1998;	98US-0085318.
XX	(LUDW-) LUDWIG INST CANCER RES.	
PA	Van Den Eynde B, Boon-Falleur T;	
XX	WPI; 2000-053076/04..	
DR	P-PSDB; AAY53809.	
XX	New isolated tumour rejection antigen RUR-1 nucleic acids, used for,	
PT	e.g. treatment of cancers -	
XX	Claim 4; Fig 5; 75pp; English.	
FS	The present sequence represents the antisense cDNA sequence of human	
CC	tumour rejection antigen RUR-1. The present sequence is the antisense	
CC	strand of a ubiquitously expressed gene. The antisense strand codes for	
CC	a polypeptide which is preferentially expressed in tumour samples and	
CC	tumour-derived cells lines. The polypeptide is unrelated to any TRAP	
CC	protein. The sequence was isolated from a renal cell carcinoma line	
CC	LB9211-ROC. The RUR-1 nucleic acids and polypeptides can be used for	
CC	diagnosis, prognosis or treatment of a disorder characterized by the	
CC	expression of a RUR-1 antisense cDNA molecule or an expression product,	
CC	such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,	
CC	melanoma, sarcoma or leukaemia.	
XX		
SQ	Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 other;	
Query Match 100.0%; Score 1382; DB 21; Length 1382;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1 CATTATGCTAACGCATTAACATCATCAGGGGTGGGAGCAGGGTCACAAAAGTGAAGTCTG 60	
Db	1 CATTATGCTAACGCATTAACATCATCAGGGGTGGGAGCAGGGTCACAAAAGTGAAGTCTG 60	
Qy	61 TCAATTCCTACTTGGAAATGAAGAGTTGAAATTAATTTAAACAGTACGGGAATGCAGACAA 120	
Db	61 TCAATTCCTACTTGGAAATGAAGAGTTGAAATTAATTTAAACAGTACGGGAATGCAGACAA 120	
Qy	121 TTTTCTCCTCTGGTGACAATATAGTGTCCAACTCTGGAGTGAATTTTAAAGATGTTTA 180	
Db	121 TTTTCTCCTCTGGTGACAATATAGTGTCCAACTCTGGAGTGAATTTTAAAGATGTTTA 180	
Qy	181 TTTAAATTTAAAGAGTGGATTTCACAGGAAAAAATAAGGAAAAAGGAAAAAACTG 240	
Db	181 TTTAAATTTAAAGAGTGGATTTCACAGGAAAAAATAAGGAAAAAGGAAAAAACTG 240	
Qy	241 AACGAAACGCCAAAGTATCAGTTGGTCTACTAACCTTTGCAAGGATACCTTTTATTT 300	
Db	241 AACGAAACGCCAAAGTATCAGTTGGTCTACTAACCTTTGCAAGGATACCTTTTATTT 300	
Qy	301 TCTTTAAGATTCCTGTGTTTATACAGATTTTAAAGTTTACTCTACTGTCACCCAAAG 360	
Db	301 TCTTTAAGATTCCTGTGTTTATACAGATTTTAAAGTTTACTCTACTGTCACCCAAAG 360	
Qy	361 TGAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCACTGCAACGAGAGTTTGT 420	
Db	361 TGAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCACTGCAACGAGAGTTTGT 420	
Qy	421 AGGGGCATCAATACACCGGAGAGTCACAGCCCTCAACCACTGAGTGTGGGGGGTAG 480	
Db	421 AGGGGCATCAATACACCGGAGAGTCACAGCCCTCAACCACTGAGTGTGGGGGGTAG 480	
Qy	481 GGATCTGCATTTCTTCATATCAACCCACACTATAGGCGACCTAAATGGGTGGCGGTGG 540	
Db	481 GGATCTGCATTTCTTCATATCAACCCACACTATAGGCGACCTAAATGGGTGGCGGTGG 540	
Qy	541 GGGAGACCGACTCACTTGAGTTTCTTGAAGGCTTCCTGGCTCCAGCCACCTAATTCGCC 600	
Db	541 GGGAGACCGACTCACTTGAGTTTCTTGAAGGCTTCCTGGCTCCAGCCACCTAATTCGCC 600	
Qy	601 CCGCTCTGGATCTGGTCTAGTCTCCGGATTCCGGTGGGCACTCCCGGGGGGTAGATGTC 660	

```
FT CDS 303..1733
FT primer_bind /*tag= a
FT 484..501
FT /*tag= b
FT /note= "binding site for primer VDE87 (see A236645)"
FT primer_bind complement (917..935)
FT /*tag= c
FT /note= "binding site for primer VDE93 (see A236646)".
XX WO958546-A1.
XX
XX 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US10424.
XX
XX 13-MAY-1998; 98US-0085318.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Van Den Eynde B, Boon-Falleur T;
XX
XX WPI: 2000-053076/04.
XX P-PSDB: AAY43611.
XX
XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,
XX e.g. treatment of cancers -
XX
XX Claim 5: Fig 6; 75pp; English.
XX
XX The present sequence represents the sense cDNA sequence of a human
XX ubiquitously expressed gene. The antisense strand of the present
XX sequence encodes tumour rejection antigen RUR-1. The antisense strand
XX codes for a polypeptide which is preferentially expressed in tumour
XX samples and tumour-derived cells lines. The polypeptide is unrelated
XX to any TRAP protein. The antisense sequence was isolated from a renal
XX cell carcinoma line LB9211-RCC. The RUR-1 nucleic acids and polypeptides
XX can be used for diagnosis, prognosis or treatment of a disorder
XX characterized by the expression of a RUR-1 antisense cDNA molecule or
XX an expression product, such as cancers, e.g. renal cell carcinoma,
XX colorectal carcinoma, melanoma, sarcoma or leukaemia.
XX
XX Sequence 2167 BP; 675 A; 440 C; 576 G; 476 T; 0 other;
XX
XX Query Match 43.1%; Score 595; DB 21; Length 2167;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-282;
XX Matches 595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 556 TTGAGTTCTTTGAAGGCTTCTGGCCTCCAGCCAGTAATTGCCCGCTCTGGATCTGG 615
DB 595 TTGAGTTCTTTGAAGGCTTCTGGCCTCCAGCCAGTAATTGCCCGCTCTGGATCTGG 536
QY 616 TCTAGTTCGGGATTCGGTGGCAGTCGCGGGGTGTAGATGTTCTGACGGCCCAAG 675
DB 535 TCTAGTTCGGGATTCGGTGGCAGTCGCGGGGTGTAGATGTTCTGACGGCCCAAG 476
QY 676 GTGCTCTGAAGCCCGCCCGTCACTCTTCAGGAAGACTTCGAAGCTGACACCTTCTTC 735
DB 475 GGTGGCTGAAGCCCGCCCGTCACTCTTCAGGAAGACTTCGAAGCTGACACCTTCTTC 416
QY 736 TCATGGATGACGACGCGCCCGCCCGGTAGAGGGGTCCCGGTTCGGGTACACAGCAG 795
DB 415 TCATGGATGACGACGCGCGCCCGCCCGGTAGAGGGGTCCCGGTTCGGGTACACAGCAG 356
QY 796 CTCCTCAGACGGGTGAGACAGGTGGCTGGACCTGGCGGTGCTGCCGCTCATCTCC 855
DB 355 CTCCTCAGACGGGTGAGACAGGTGGCTGGACCTGGCGGTGCTGCCGCTCATCTCC 296
QY 856 GCTGCGCCCGCTCAGCTGCTGCTTCGCTCGGGAGGACCTCCGCTGTCCAGCGGC 915
DB 295 GCTGCGCCCGCTCAGCTGCTGCTTCGCTCGGGAGGACCTCCGCTGTCCAGCGGC 236
QY 916 CTCACCGCACCCAGGCGCGGGATCGCTCTCTGAAACGAAGAGAACTGACGAATCCAC 975
DB 916 CTCACCGCACCCAGGCGCGGGATCGCTCTCTGAAACGAAGAGAACTGACGAATCCAC 975
```

```
Db 235 CTCACCGCACCCAGGCGCGGGATCGCTCTGAAACGAAGAGAACTGACGAATCCAC 176
QY 976 AGGTGAAGAGAAGTAACGCGCTAGCGCTCCACCCAGAGGAGACACTAGGAGC 1035
Db 175 AGGTGAAGAGAAGTAACGCGCTAGCGCTCCACCCAGAGGAGACACTAGGAGC 116
QY 1036 TTGCAGGACTCGGAGTAGACGCTCAAGTTTTCACCGCTGGCGTGACAGCCCAATCAGGAC 1095
Db 115 TTGCAGGACTCGGAGTAGACGCTCAAGTTTTCACCGCTGGCGTGACAGCCCAATCAGGAC 56
QY 1096 CCGCAGTGGCGGCACACACAGGTTTTCACCTGCTACGGGCGAGAAATCAAGGTGGAC 1150
Db 55 CCGCAGTGGCGGCACACACAGGTTTTCACCTGCTACGGGCGAGAAATCAAGGTGGAC 1
```

## RESULT 3

AAI58192/c

ID AAI58192 standard; cDNA; 1956 BP.

XX AC AAI58192;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 395.

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;  
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX P-PSDB: AAM39036.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 395; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with nontropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 1956 BP; 540 A; 435 C; 541 G; 435 T; 5 other;  
Query Match 36.0%; Score 498; DB 22; Length 1956;  
Best Local Similarity 100.0%; Pred. No. 8.8e-235; Mismatches 0; Indels 0; Gaps 0;  
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 556 TTGAGTTCTTGAAGGCTTCTCGCTCCAGCCACGTAATTCGCCCGCTCTGGATCTGG 615  
DB 593 TTGAGTTCTTGAAGGCTTCTCGCTCCAGCCACGTAATTCGCCCGCTCTGGATCTGG 594  
QY 616 TCTAGCTTCGGATTCGGTGGCAGTCCGGGGGTGTAGATGTTCTGAGGCCCCAAAG 675  
DB 533 TCTAGCTTCGGATTCGGTGGCAGTCCGGGGGTGTAGATGTTCTGAGGCCCCAAAG 474  
QY 676 GGTGCTTGAACGCGCGGTCACCTCTTCAGGAAGACTTCGAAGCTGGACACTTCTTC 735  
DB 473 GGTGCTTGAACGCGCGGTCACCTCTTCAGGAAGACTTCGAAGCTGGACACTTCTTC 414  
QY 736 TCATGATGACGACGCGCGGCCCGGTAGAGGGGTCCCGGTTGGGGTACACAAACGACG 795  
DB 413 TCATGATGACGACGCGCGGCCCGGTAGAGGGGTCCCGGTTGGGGTACACAAACGACG 354  
QY 796 CTCTTCACGACGGGTGAGACAGGTGGCTGGACCTGGCGGTGTGCGGTCTATCTTCCCC 855  
DB 353 CTCTTCACGACGGGTGAGACAGGTGGCTGGACCTGGCGGTGTGCGGTCTATCTTCCCC 294  
QY 856 GCTGGCGCGCGCTCAGCTCGCTGCTCGGTGGGAGGACCTCCGCTCTCCACGGGC 915  
DB 293 GCTGGCGCGCGCTCAGCTCGCTGCTCGGTGGGAGGACCTCCGCTCTCCACGGGC 234  
QY 916 CTCACCGCACCCAGGCGCGGGATCGCTCTCTGAACAGCAACGAGAACTGACGAATCCAC 975  
DB 233 CTCACCGCACCCAGGCGCGGGATCGCTCTCTGAACAGCAACGAGAACTGACGAATCCAC 174  
QY 976 AGGTGAAGAAGATTAACGGCCGTGGCGCTAGCGCTCCACCCAGAGAGACATAGAGC 1035  
DB 173 AGGTGAAGAAGATTAACGGCCGTGGCGCTAGCGCTCCACCCAGAGAGACATAGAGC 114  
QY 1036 TTGCAGGACTCGGAGTAG 1053  
DB 113 TTGCAGGACTCGGAGTAG 96  
RESULT 4  
AAZ36649  
ID AAZ36649 standard; cDNA; 246 BP.  
XX  
XX AAZ36649;  
AC  
XX  
XX 22-FEB-2000 (first entry)  
DE EST AA863443 which overlaps the RUR-1 antisense cDNA sequence.  
XX  
XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia; EST; expressed sequence tag; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO9558546-A1.  
PN  
XX  
XX 18-NOV-1999.  
PD  
XX  
XX 13-MAY-1999; 99WO-US10424.  
PF  
XX  
XX 13-MAY-1998; 98US-0085318.  
PR

XX (LUDW-) LUDWIG INST CANCER RES.  
PA Van Den Eynde B, Boon-Falleur T;  
PI  
XX  
XX WPI; 2000-053076/04.  
DR  
XX  
XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,  
PT e.g. treatment of cancers -  
PT  
XX  
XX Claim 8; Page 71; 75pp; English.  
PS  
XX  
XX The present sequence represents an expressed sequence tag (EST)  
CC which overlaps the antisense cDNA sequence of human tumour rejection  
CC antigen RUR-1. The RUR-1 antisense sequence is the antisense strand  
CC of a ubiquitously expressed gene. The antisense strand codes for  
CC a polypeptide which is preferentially expressed in tumour samples and  
CC tumour-derived cells lines. The polypeptide is unrelated to any TRAP  
CC protein. The RUR-1 sequence was isolated from a renal cell carcinoma  
CC line L89211-RCC. The RUR-1 nucleic acids and polypeptides can be used  
CC for diagnosis, prognosis or treatment of a disorder characterized by the  
CC expression of a RUR-1 antisense cDNA molecule or an expression product,  
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,  
CC melanoma, sarcoma or leukaemia.  
CC  
CC note: although the present sequence is mentioned in claim 8, it is not  
CC specifically claimed. The fragments of AAZ36643-44 which DO NOT contain  
CC the present sequence are claimed.  
XX  
XX Sequence 246 BP; 63 A; 56 C; 44 G; 73 T; 0 other;  
SQ  
Query Match 17.4%; Score 241; DB 21; Length 246;  
Best Local Similarity 100.0%; Pred. No. 3e-108; Mismatches 0; Indels 0; Gaps 0;  
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 273 TTAACCTTTCAGGATACCTTTTATTTCTTTAAAGTTCCTGTTTATACACAGATT 332  
DB 6 TAACTTTTCAAGGATACCTTTTATTTCTTTAAAGTTCCTGTTTATACACAGATT 65  
QY 333 TTAAGTTTACTCTCTACTGCTGACCCCAAGTGAAATTCCTTCCAGTCACAGTCAACCT 392  
DB 66 TTAAGTTTACTCTCTACTGCTGACCCCAAGTGAAATTCCTTCCAGTCACAGTCAACCT 125  
QY 393 CTACCCCCCACTGCAACGAGAGTTCCTTGGGGGCATCAATCACACGAGAGTCACAGCC 452  
DB 126 CTACCCCCCACTGCAACGAGAGTTCCTTGGGGGCATCAATCACACGAGAGTCACAGCC 185  
QY 453 CCTCAACCACTGAGGTGCTGGGGGGTAGGATCTGCAATTCCTCATATCAACCCCACT 512  
DB 186 CCTCAACCACTGAGGTGCTGGGGGGTAGGATCTGCAATTCCTCATATCAACCCCACT 245  
QY 513 A 513  
DB 246 A 246  
RESULT 5  
AAI59978/c  
ID AAI59978 standard; cDNA; 492 BP.  
XX  
XX AAI59978;  
AC  
XX  
XX 22-OCT-2001 (first entry)  
DT  
XX  
XX Human polynucleotide SEQ ID NO 3967.  
DE  
XX  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemofactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
XX Homo sapiens.  
OS

```
XX WO200153312-A1.
PN
XX
XX
PD 26-JUL-2001.
XX
XX PF 26-DEC-2000; 2000WO-US34263.
XX
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR
DR P-PSDB; AAM40822.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 3967; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 492 BP; 145 A; 111 C; 132 G; 100 T; 4 other;
SQ
Query Match 10.1%; Score 139; DB 22; Length 492;
Best Local Similarity 100.0%; Pred. No. 5e-58;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 556 TTGAGTTTCTTGAAGGCTTCCTGGCTCCAGCCACGTAATTCGCCCGCTCTGGATCTGG 615
Db |||||||
191 TTGAGTTTCTTGAAGGCTTCCTGGCTCCAGCCACGTAATTCGCCCGCTCTGGATCTGG 132

QY 616 TCTAGCTTCGCGATTCGGTGGCCAGTCCCGGGGTGTAGATGTTCTGCAGGCCCCCAAG 675
Db |||||||
131 TCTAGCTTCGCGATTCGGTGGCCAGTCCCGGGGTGTAGATGTTCTGCAGGCCCCCAAG 72

QY 676 GGTGCTGAACGCCGCCCG 694
Db |||||||
71 GGTGCTGAACGCCGCCCG 53

RESULT 6
ABN38745
ID ABN38745 standard; DNA; 60 BP.
XX
AC ABN38745;
XX
XX 15-JUL-2002 (first entry)
DT
```

```
XX Human spliced transcript detection oligonucleotide SEQ ID NO:11493.
DE
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
XX Homo sapiens.
OS
XX WO200210449-A2.
PN
XX 07-FEB-2002.
PD
XX 20-JUL-2001; 2001WO-IB01903.
XX
XX 28-JUL-2000; 2000US-221607P.
XX 02-MAY-2001; 2001US-287724P.
PR
XX (COMP-) COMPUGEN INC.
XX
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
PI
XX WPI: 2002-257383/30.
XX
XX New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
XX Example 1; SEQ ID 11493; 47pp; English.
XX
XX The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 BP; 23 A; 15 C; 15 G; 7 T; 0 other;
SQ
Query Match 4.3%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.9e-19;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 938 ATCGCCTCTCTGAACGACGAGAACTGACGAATCCACAGGTGAAGAAGAGTAACGGCC 997
Db |||||||
1 ATCGCCTCTCTGAACGACGAGAACTGACGAATCCACAGGTGAAGAAGAGTAACGGCC 60

RESULT 7
AAC99033
ID AAC99033 standard; cDNA; 2116 BP.
XX
AC AAC99033;
XX
XX 09-MAR-2001 (first entry)
DT
```

XX DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:261.

XX DE Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

KW detection; diagnosis; identification; relaxant; cytostatic; neuroprotective;

KW neoplastic; immunomodulatory; relaxant; gene therapy; chromosome mapping;

KW antineoplastic; relaxant; gene therapy; chromosome mapping;

KW linkage analysis; tissue identification; tissue typing; forensic;

KW neural; immune system; muscular; reproductive; gastrointestinal;

KW pulmonary; cardiovascular; renal; proliferative; ss.

XX OS Homo sapiens.

XX PN WO200055320-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05989.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX P-PSDB; AAB54466.

XX WPI; 2000-579444/54.

XX DR New nucleic acid that is a pancreatic cancer antigen for preventing,

XX PT treating, or ameliorating a medical condition, particular pancreatic

XX PT cancer, or for use in assays for diagnosing a pathological condition -

XX PS Claim 1; Page 699-700; 1379pp; English.

XX CC AAC98773 to AAC99231 encode the human pancreatic cancer associated

XX CC proteins, called pancreatic cancer antigens, given in AAB54008 to

XX CC AAB54466. The human pancreatic cancer antigens have cytostatic,

XX CC neuroprotective, nontoxic, immunomodulatory, relaxant, contractile,

XX CC gynaecological, cardiac and anti-inflammatory activities, and can be used

XX CC in gene therapy. The polynucleotide and proteins can be used for

XX CC preventing, treating, or ameliorating a medical condition or in assays

XX CC for diagnosing a pathological condition or a susceptibility to one in a

XX CC subject. Binding partners to the proteins and the activity of the

XX CC proteins can be identified. The pancreatic cancer antigens can be used to

XX CC detect, treat or prevent pancreatic disorders, especially cancer.

XX CC Agonists and antagonists to the antigens can be screened for. The

XX CC pancreatic cancer antigen polynucleotides can be used to design nucleic

XX CC acid hybridisation probes that can be used in chromosome mapping, linkage

XX CC analysis, tissue identification and/or typing and a variety of forensic

XX CC and diagnostic methods. The proteins can be used to generate antibodies

XX CC which are used to purify, detect and target the polypeptides, including

XX CC both in vivo and in vitro diagnostic and therapeutic methods. The

XX CC proteins can be used to treat or prevent neural, immune system, muscular,

XX CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or

XX CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent

XX CC sequences used in the exemplification of the present invention.

XX SQ Sequence 2116 BP; 606 A; 485 C; 390 G; 629 T; 6 other;

Query Match 2.0%; Score 28; DB 21; Length 2116;

Best Local Similarity 100.0%; Pred. No. 0.0022;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1225 CGCACACACACACACAAATATG 1252

Db 204 CGCACACACACACACAAATATG 231

RESULT 8

AAS00847

ID AAS00847 standard; cDNA; 2116 BP.

XX AAS00847;

XX 04-JUL-2001 (first entry)

XX Human cDNA clone HIBL55 encoding cancer related protein 21.

DE Human; cancer related protein; HIBL55; food additive;

XX preservative; immunogen; antibody; bone cancer; adrenal cancer;

KW bone marrow cancer; breast cancer; gastrointestinal cancer;

KW liver cancer; lung cancer; urogenital cancer; immune disorder;

KW Addison's disease; allergy; autoimmune haemolytic anaemia;

KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;

KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;

KW acquired immunodeficiency syndrome; AIDS; cardiovascular disorder;

KW myocardial ischaemia; wound healing; neurological disorder;

KW Parkinson's disease; Alzheimer's disease; cerebral anoxia; epilepsy;

KW viral infection; bacterial infection; fungal infection;

KW parasitic infection; agonist; antagonist; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT CDS 342..482

FT /\*tag= a

FT /product= "Cancer related protein 21"

XX WO200118014-A1.

XX PD 15-MAR-2001.

XX PF 30-AUG-2000; 2000WO-US23794.

XX PR 03-SEP-1999; 99US-0152296.

XX PR 06-OCT-1999; 99US-0158003.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Roschke V;

XX WPI; 2001-235186/24.

XX P-PSDB; AAU00886.

XX Twenty nine nucleic acid molecules encoding human cancer associated

XX proteins, useful in the prevention, treatment and diagnosis of cancer,

XX immune disorders, cardiovascular disorders and neurological diseases -

XX Disclosure; Page 379-380; 427pp; English.

XX The sequence encodes a novel Human cancer related protein. The

XX polynucleotides and polypeptides are useful for preventing,

XX treating or ameliorating a medical condition in e.g. humans,

XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The

XX polypeptides can also be used as a food additive or preservative to

XX increase or decrease storage capabilities. The polynucleotide are

XX useful for chromosome identification. The nucleic acids, protein,

XX antibodies, agonists and antagonists are useful in the diagnosis,

XX treatment and prevention of cancer (e.g. cancers of the adrenal gland,

XX bone, bone marrow, breast, gastrointestinal tract, liver, lung, or

XX urogenital), immune disorders (e.g. Addison's disease, allergies,

XX autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,

XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative

XX colitis, acquired immunodeficiency syndrome, AIDS), cardiovascular

XX disorders (e.g. as myocardial ischaemias, wound healing, neurological

XX diseases (e.g. Parkinson's disease, Alzheimer's disease, cerebral anoxia

XX and epilepsy) and infectious diseases such as viral, bacterial, fungal

XX and parasitic infections. Numerous examples of each type of disorder are

XX given in the specification.

XX SQ Sequence 2116 BP; 606 A; 485 C; 390 G; 629 T; 6 other;

Query Match 2.0%; Score 28; DB 22; Length 2116;

Best Local Similarity 100.0%; Pred. No. 0.0022;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1225 CGCACACACACACACACACACAAATATG 1252  
          |||||  
Db 204 CGCACACACACACACACACAAATATG 231

RESULT 9  
ID ABA07127 standard; DNA; 2219 BP.  
XX ABA07127;  
AC ABA07127;  
XX  
XX 14-JAN-2002 (first entry)  
XX  
XX Human pancreatic cancer related genomic DNA, SEQ ID NO: 446.  
XX  
XX Human: cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;  
KW antihormone; antiulcer; thyroid-active; gene therapy; antisense therapy;  
KW pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;  
KW diabetes; endocrine disorder; acromegaly; hyperthyroidism;  
KW gastrointestinal disorder; Crohn's disease; duodenal ulcer; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200155206-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01353.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.

```
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251969.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-457717/49.
XX
XX Isolated pancreatic cancer polypeptide for treating, preventing and/ or
PT prognosing disorders related to the pancreas including pancreatic
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Disclosure; SEQ ID NO 446; 537pp; English.
XX
XX The invention relates to an isolated polypeptide comprising an amino
CC acid sequence at least 90% identical to 188 amino acid sequences fully
CC defined in the specification and encoded by 188 cDNA clones fully
CC defined in the specification. The invention also relates to a fragment
CC having biological activity, a domain, an epitope, full length protein,
CC variant, allelic variant or a species homologue of the fully defined
CC sequence. The polynucleotide and polypeptide are useful for treating,
CC preventing and/or prognosing disorders related to the pancreas including
CC pancreatic cancer, pancreatitis, diabetes, endocrine disorders such as
CC acromegaly or hyperthyroidism, and gastrointestinal disorders such as
CC Crohn's disease and duodenal ulcers. The present sequence encodes a
CC pancreatic cancer-related polypeptide of the invention.
XX
XX Sequence 2219 BP; 618 A; 507 C; 409 G; 685 T; 0 other;
SQ
Query Match 2.0%; Score 28; DB 22; Length 2219;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1225 CGCACACACACACACACAATATG 1252
Db 462 CGCACACACACACACACAATATG 489
|||||
RESULT 10
AAK89341
ID AAK89341 standard; DNA; 2219 BP.
XX
AC AAK89341;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 2017.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01324.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
```

PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0246174.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
PI  
XX

DR WPI; 2001-502630/55.  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
PS Disclosure; SEQ ID NO 2917; 986pp; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention.  
XX  
SQ Sequence 2219 BP; 618 A; 507 C; 409 G; 685 T; 0 other;  
Query Match 2.0%; Score 28; DB 22; Length 2219;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1225 CGCACACACACACACACACAAATATG 1252  
Db 462 CGCACACACACACACACACAAATATG 489  
RESULT 11  
ABK83460/c  
ID ABK83460 standard; cDNA; 165199 BP.  
XX  
AC ABK83460;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #31.  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX Homo sapiens.  
OS  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
XX  
PR 03-OCT-2000; 2000US-237189P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX  
XX WPI; 2002-435328/46.  
DR  
XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity -  
XX  
PS Claim 1; SEQ ID No 31; 114pp; English.  
XX  
CC The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated

GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC periodontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 other;

Query Match 2.0%; Score 28; DB 24; Length 165199;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1221 AGGACGCACACACACACACACAA 1248  
|||||  
DB 17664 AGGACGCACACACACACACACAA 17637

RESULT 12  
AAI82036/c  
ID AAI82036 standard; cDNA; 383 BP.

XX AAI82036;

XX 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 2096.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

XX WO200164835-A2.

XX 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.  
DR P-PSDB; AAO02105.  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX Claim 1; SEQ ID NO 2096; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 383 BP; 95 A; 38 C; 84 G; 165 T; 1 other;

Query Match 1.9%; Score 26; DB 22; Length 383;

Best Local Similarity 100.0%; Pred. No. 0.021;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1227 CACACACACACACACACAAATATG 1252

|||||  
DB 228 CACACACACACACACACAAATATG 203

RESULT 13

AAO05327/c

ID AAO05327 standard; cDNA; 1747 BP.

XX AAO05327;

XX 17-JUL-2001 (first entry)

XX Human secreted protein-encoding gene 28 cDNA clone HNGPM78, SEQ ID NO:38.

XX Human; secreted protein; proliferative disorder; cancer; tumour;  
KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KW cardiovascular disorder; angioecnic disorder; kidney disorder;  
KW gastrointestinal disorder; pregnancy-related disorder;  
KW endocrine disorder; infection; wound healing; vulnery;  
KW cell culture; chemotaxis; food additive; gene therapy;  
KW binding partner identification; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 307..384

FT /tag= a  
FT /product= "Human secreted protein"

FT sig\_peptide 307..366

FT /tag= b

FT mat\_peptide 367..381

FT /tag= c  
FT /product= "Human mature secreted protein"

XX WO200134626-A1.

XX 17-MAY-2001.

PD

```
XX 01-NOV-2000; 2000WO-US30045.
XX PF
XX XX
XX 05-NOV-1999; 99US-0163581.
XX PR
XX 30-JUN-2000; 2000US-0215133.
XX PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX XX
XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;
XX PI
XX WPI; 2001-308778/32.
XX DR
XX P-PSDB; AAE01463.
XX DR
XX New nucleic acid molecules encoding 28 human secreted proteins for
XX PT diagnosing, preventing, treating or ameliorating medical conditions and
XX PT used as food additives or preservatives.
XX XX
XX Claim 1; Page 443-444; 562pp; English.
XX PS
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted
XX CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.
XX CC AAE01514-AAE01544 represent human secreted protein fragments or variants.
XX CC The genes and their secreted proteins are useful for preventing,
XX CC treating or ameliorating medical conditions, e.g., by protein or gene
XX CC therapy. Pathological conditions can be diagnosed by determining the
XX CC amount of the new protein in a sample or by determining the presence of
XX CC mutations in the new genes. Specific uses are described for each of the
XX CC 28 genes, based on the tissues in which they are most highly expressed,
XX CC and include developing products for the diagnosis or treatment of
XX CC proliferative disorders, cancer, tumours, foetal and developmental
XX CC abnormalities, haematopoietic disorders, diseases of the immune system,
XX CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
XX CC allergies, neurological disorders (e.g., Alzheimer's disease,
XX CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
XX CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
XX CC cardiovascular disorders, angiotensin disorders, kidney disorders,
XX CC gastrointestinal disorders, pregnancy-related disorders, endocrine
XX CC disorders, and infections. The proteins can also be used to aid wound
XX CC healing and epithelial cell proliferation, to prevent skin aging due to
XX CC sunburn, to maintain organs before transplantation, for supporting cell
XX CC culture of primary tissues, to regenerate tissues, to identify their
XX CC cognate ligands or binding partners, and in chemotaxis, and can be used
XX CC as a food additive or preservative to modify storage properties.
XX CC Antibodies specific for a protein of the invention can be used in
XX CC alleviating symptoms associated with the disorders mentioned above, and
XX CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
XX CC immunosorbent assay (ELISA). The present sequence represents a human
XX CC secreted protein-encoding cDNA of the invention.
XX XX
XX Sequence 1747 BP; 407 A; 481 C; 476 G; 383 T; 0 other;
XX SQ
Query Match 1.9%; Score 26; DB 22; Length 1747;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1227 CACACACACACACACACAAATATG 1252
Db 1208 CACACACACACACACACAAATATG 1183
RESULT 14
AAK83864
ID AAK83864 standard; DNA; 16584 BP.
XX AC
XX AAK83864;
XX XX
XX 07-NOV-2001 (first entry)
XX DT
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38676.
XX DE
XX Human; immune: haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX XX
```

```
OS Homo sapiens.
XX XX
XX WO200157182-A2.
XX XX
XX 09-AUG-2001.
XX PF
XX 17-JAN-2001; 2001WO-US01354.
XX XX
XX 31-JAN-2000; 2000US-0179065.
XX PR
XX 04-FEB-2000; 2000US-0180628.
XX PR
XX 24-FEB-2000; 2000US-0184664.
XX PR
XX 02-MAR-2000; 2000US-0186350.
XX PR
XX 16-MAR-2000; 2000US-0189874.
XX PR
XX 17-MAR-2000; 2000US-0190076.
XX PR
XX 18-APR-2000; 2000US-0198123.
XX PR
XX 19-MAY-2000; 2000US-0205515.
XX PR
XX 07-JUN-2000; 2000US-0209467.
XX PR
XX 28-JUN-2000; 2000US-0214886.
XX PR
XX 30-JUN-2000; 2000US-0215135.
XX PR
XX 07-JUL-2000; 2000US-0216647.
XX PR
XX 07-JUL-2000; 2000US-0216880.
XX PR
XX 11-JUL-2000; 2000US-0217487.
XX PR
XX 11-JUL-2000; 2000US-0217496.
XX PR
XX 14-JUL-2000; 2000US-0218290.
XX PR
XX 26-JUL-2000; 2000US-0220963.
XX PR
XX 26-JUL-2000; 2000US-0220964.
XX PR
XX 14-AUG-2000; 2000US-0224518.
XX PR
XX 14-AUG-2000; 2000US-0224519.
XX PR
XX 14-AUG-2000; 2000US-0225213.
XX PR
XX 14-AUG-2000; 2000US-0225214.
XX PR
XX 14-AUG-2000; 2000US-0225266.
XX PR
XX 14-AUG-2000; 2000US-0225267.
XX PR
XX 14-AUG-2000; 2000US-0225268.
XX PR
XX 14-AUG-2000; 2000US-0225270.
XX PR
XX 14-AUG-2000; 2000US-0225447.
XX PR
XX 14-AUG-2000; 2000US-0225757.
XX PR
XX 14-AUG-2000; 2000US-0225758.
XX PR
XX 14-AUG-2000; 2000US-0225759.
XX PR
XX 18-AUG-2000; 2000US-0226279.
XX PR
XX 22-AUG-2000; 2000US-0226681.
XX PR
XX 22-AUG-2000; 2000US-0226868.
XX PR
XX 22-AUG-2000; 2000US-0227182.
XX PR
XX 23-AUG-2000; 2000US-0227009.
XX PR
XX 30-AUG-2000; 2000US-0228924.
XX PR
XX 01-SEP-2000; 2000US-0229287.
XX PR
XX 01-SEP-2000; 2000US-0229343.
XX PR
XX 01-SEP-2000; 2000US-0229344.
XX PR
XX 01-SEP-2000; 2000US-0229345.
XX PR
XX 05-SEP-2000; 2000US-0229509.
XX PR
XX 05-SEP-2000; 2000US-0229513.
XX PR
XX 06-SEP-2000; 2000US-0230437.
XX PR
XX 06-SEP-2000; 2000US-0230438.
XX PR
XX 08-SEP-2000; 2000US-0231242.
XX PR
XX 08-SEP-2000; 2000US-0231243.
XX PR
XX 08-SEP-2000; 2000US-0231244.
XX PR
XX 08-SEP-2000; 2000US-0231413.
XX PR
XX 08-SEP-2000; 2000US-0231414.
XX PR
XX 08-SEP-2000; 2000US-0232080.
XX PR
XX 08-SEP-2000; 2000US-0232081.
XX PR
XX 12-SEP-2000; 2000US-0231968.
XX PR
XX 14-SEP-2000; 2000US-0232397.
XX PR
XX 14-SEP-2000; 2000US-0232398.
XX PR
XX 14-SEP-2000; 2000US-0232399.
XX PR
XX 14-SEP-2000; 2000US-0232400.
XX PR
XX 14-SEP-2000; 2000US-0232401.
XX PR
XX 14-SEP-2000; 2000US-0233063.
XX PR
XX 14-SEP-2000; 2000US-0233064.
XX PR
XX 14-SEP-2000; 2000US-0233065.
XX PR
XX 21-SEP-2000; 2000US-0234223.
XX PR
XX 21-SEP-2000; 2000US-0234224.
XX PR
XX 25-SEP-2000; 2000US-0234997.
XX PR
XX 25-SEP-2000; 2000US-0234998.
XX PR
XX 26-SEP-2000; 2000US-0235484.
XX PR
```

PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 20-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241285.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 05-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX Disclosure; SEQ ID NO 38676; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
XX represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 16584 BP; 3944 A; 4426 C; 4384 G; 3828 T; 2 other;  
XX  
XX Query Match 1.9%; Score 26; DB 22; Length 16584;  
XX Best Local Similarity 100.0%; Pred. No. 0.021;  
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1227 CACACACACACACACACAAATATG 1252  
DB 14734 CACACACACACACACACAAATATG 14759  
|||||  
RESULT 15  
AAZ36647  
ID AAZ36647 standard; cDNA; 25 BP.  
XX  
XX AAZ36647;  
XX  
XX DT 22-FEB-2000 (first entry)  
XX  
XX PCR primer VDE119 used to amplify RUR-1 antisense cDNA sequence.  
XX  
XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
XX renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
XX leukaemia; PCR primer; ss.  
XX  
XX Synthetic.  
XX Homo sapiens.  
XX  
XX WO9958546-A1.  
XX  
XX PD 18-NOV-1999.  
XX  
XX PF 13-MAY-1999; 99WO-US10424.  
XX  
XX PR 13-MAY-1998; 98US-0085318.  
XX  
XX (LUDW-) LUDWIG INST CANCER RES.  
XX  
XX Van Den Eynde B, Boon-Falleur T;  
XX  
XX WPI; 2000-053076/04.  
XX  
XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,  
XX e.g. treatment of cancers -  
XX  
XX



**THIS PAGE BLANK (USP10)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:37:54 ; Search time 46 Seconds  
(without alignments)  
166.672 Million cell updates/sec

Title: US-09-674-593-8  
Perfect score: 25  
Sequence: 1 taaatgggtggcggtggggagac 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.8	79.2	12141	4	US-09-488-671-10
C 2	19.2	76.8	1000	4	US-09-641-638-631
C 3	19.2	76.8	1000	4	US-09-641-638-632
C 4	19.2	76.8	1000	4	US-09-641-638-633
C 5	19.2	76.8	1000	4	US-09-641-638-634
C 6	19.2	76.8	20674	4	US-09-641-638-651
C 7	18.6	74.4	2605	2	US-08-680-395-4
C 8	17.6	70.4	3211	1	US-07-674-287B-1
C 9	17.6	70.4	3211	2	US-08-436-900A-1
C 10	17	68.0	50	1	US-08-171-389-505
C 11	17	68.0	50	1	US-08-123-936-505
C 12	17	68.0	50	2	US-08-475-228A-505
C 13	17	68.0	50	3	US-08-482-080A-505
C 14	17	68.0	50	4	US-09-354-947-505
C 15	17	68.0	50	5	PCT-US93-12388-505
C 16	16.8	67.2	1687	3	US-08-656-984A-33
C 17	16.6	66.4	1741	3	US-08-795-430-12
C 18	16.6	66.4	1741	4	US-09-355-700-12
C 19	16.6	66.4	2230	1	US-08-217-327-5
C 20	16.6	66.4	2255	4	US-08-871-572B-3
C 21	16.6	66.4	4105	4	US-08-121-446-1
C 22	16.6	66.4	37950	4	US-09-338-907-183
C 23	16.6	66.4	37950	4	US-09-218-207-183
C 24	16.4	65.6	1524	4	US-09-111-730-4
C 25	16.4	65.6	2469	4	US-09-111-730-6
C 26	16.2	64.8	30	1	US-08-467-126-1
C 27	16.2	64.8	30	2	US-08-476-712-3
C 1	19.8	79.2	12141	4	US-09-488-671-10
C 2	19.2	76.8	1000	4	US-09-641-638-631
C 3	19.2	76.8	1000	4	US-09-641-638-632
C 4	19.2	76.8	1000	4	US-09-641-638-633
C 5	19.2	76.8	1000	4	US-09-641-638-634
C 6	19.2	76.8	20674	4	US-09-641-638-651
C 7	18.6	74.4	2605	2	US-08-680-395-4
C 8	17.6	70.4	3211	1	US-07-674-287B-1
C 9	17.6	70.4	3211	2	US-08-436-900A-1
C 10	17	68.0	50	1	US-08-171-389-505
C 11	17	68.0	50	1	US-08-123-936-505
C 12	17	68.0	50	2	US-08-475-228A-505
C 13	17	68.0	50	3	US-08-482-080A-505
C 14	17	68.0	50	4	US-09-354-947-505
C 15	17	68.0	50	5	PCT-US93-12388-505
C 16	16.8	67.2	1687	3	US-08-656-984A-33
C 17	16.6	66.4	1741	3	US-08-795-430-12
C 18	16.6	66.4	1741	4	US-09-355-700-12
C 19	16.6	66.4	2230	1	US-08-217-327-5
C 20	16.6	66.4	2255	4	US-08-871-572B-3
C 21	16.6	66.4	4105	4	US-08-121-446-1
C 22	16.6	66.4	37950	4	US-09-338-907-183
C 23	16.6	66.4	37950	4	US-09-218-207-183
C 24	16.4	65.6	1524	4	US-09-111-730-4
C 25	16.4	65.6	2469	4	US-09-111-730-6
C 26	16.2	64.8	30	1	US-08-467-126-1
C 27	16.2	64.8	30	2	US-08-476-712-3

28	16.2	64.8	30	4	US-09-411-291-3	Sequence 3, Appl1
C 29	16.2	64.8	728	4	US-08-998-416-274	Sequence 274, App
C 30	16.2	64.8	783	4	US-09-149-476-270	Sequence 270, App
C 31	16.2	64.8	1028	4	US-09-071-035-103	Sequence 103, App
C 32	16.2	64.8	1122	4	US-09-071-035-101	Sequence 101, App
C 33	16.2	64.8	3157	2	US-08-693-174-5	Sequence 5, Appl1
C 34	16.2	64.8	3157	4	US-08-253-738-5	Sequence 5, Appl1
C 35	16.2	64.8	3460	2	US-08-751-305-1	Sequence 1, Appl1
C 36	16.2	64.8	6370	4	US-09-245-041-12	Sequence 12, Appl1
C 37	16.2	64.8	7812	4	US-09-368-590-1	Sequence 1, Appl1
C 38	16.2	64.8	8589	4	US-09-245-041-14	Sequence 14, Appl1
C 39	16.2	64.8	8827	4	US-09-245-041-1	Sequence 1, Appl1
C 40	16.2	64.8	29629	4	US-09-729-995-3	Sequence 3, Appl1
C 41	16.2	64.8	53526	3	US-08-658-136-2	Sequence 2, Appl1
C 42	16.2	64.8	53577	3	US-08-658-136-1	Sequence 1, Appl1
C 43	16	64.0	1254	3	US-08-815-469-3	Sequence 3, Appl1
C 44	16	64.0	1634	4	US-08-928-069-11	Sequence 11, Appl1
C 45	16	64.0	1634	4	US-08-828-683A-9	Sequence 9, Appl1

## ALIGNMENTS

RESULT 1  
US-09-488-671-10/c  
; Sequence 10, Application US/09488671A  
; Patent No. 6187545  
; GENERAL INFORMATION:  
; APPLICANT: Robert McKay  
; APPLICANT: Madeline M. Butler  
; APPLICANT: Jacqueline Wyatt  
; APPLICANT: Lex M. Cowsett  
; TITLE OF INVENTION: ANTISENSE MODULATION OF PEPCK-CYTOSOLIC EXPRESSION  
; FILE REFERENCE: RTS-0123  
; CURRENT APPLICATION NUMBER: US/09/488,671A  
; CURRENT FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 177  
; SEQ ID NO 10  
; LENGTH: 12141  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (5895)...(6118)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (6440)...(6621)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7206)...(7409)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7682)...(7869)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (8444)...(8606)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (9418)...(9642)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (9735)...(9866)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10502)...(10597)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10701)...(11155)  
US-09-488-671-10

Query Match 79.2%; Score 19.8; DB 4; Length 12141;  
Best Local Similarity 91.3%; Pred. No. 9;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 TAAATGGTGGCGGTGGGGGAG 23
||||||| ||| |||||
Db 8237 TAAATGGGAGGAGGTGGGGAG 8215

RESULT 2
US-09-641-638-631
; Sequence 631, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 631
; LENGTH: 1000
; TYPE: DNA
; ORGANISM: Homo Sapiens
; NAME/KEY: allele
; LOCATION: 501
; OTHER INFORMATION: 10-507-170 : polymorphic base A or G
; NAME/KEY: misc.binding
; LOCATION: 481..500
; OTHER INFORMATION: 10-507-170.misl, potential
; NAME/KEY: misc.binding
; LOCATION: 502..521
; OTHER INFORMATION: 10-507-170.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 332..350
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 739..758
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc.binding
; LOCATION: 489..513
; OTHER INFORMATION: 10-507-170 potential probe
; US-09-641-638-631

Query Match 76.8%; Score 19.2; DB 4; Length 1000;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATGGTGGCGGTGGGGGAGAC 25
||||| ||||| |||||
Db 53 AAATGCTGGCGGGGGGGGAGAC 76

RESULT 4
US-09-641-638-633
; Sequence 633, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
```

SEQ ID NO 633  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
NAME/KEY: misc\_binding  
LOCATION: 481..500  
OTHER INFORMATION: 10-507-353.misl, potential  
NAME/KEY: misc\_binding  
LOCATION: 502..521  
OTHER INFORMATION: 10-507-353.mis2, potential complement  
NAME/KEY: primer\_bind  
LOCATION: 149..167  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer\_bind  
LOCATION: 556..575  
OTHER INFORMATION: downstream amplification primer, complement  
NAME/KEY: misc\_binding  
LOCATION: 489..513  
OTHER INFORMATION: 10-507-353 potential probe  
US-09-641-638-633

Query Match 76.8%; Score 19.2; DB 4; Length 1000;  
Best Local Similarity 87.5%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAATGGTGGCGGTGGGGGAGAC 25  
||||| ||||| ||||| ||||| |||||  
Db 21 AAATGGTGGCGGGGGGGGAGAC 44

## RESULT 5

US-09-641-638-634  
Sequence 634, Application US/09641638  
Patent No. 6432648  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
FILE REFERENCE: GENSET.051CP1  
CURRENT APPLICATION NUMBER: US/09/641.638  
PRIOR FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: US 09/502,330  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: US 60/133,200  
PRIOR FILING DATE: 1999-05-07  
PRIOR APPLICATION NUMBER: US 09/275,267  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: US 60/119,917  
PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: Patent.pm  
SEQ ID NO 634  
LENGTH: 1000  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: allele  
LOCATION: 501  
OTHER INFORMATION: 10-507-364 : polymorphic base C or T  
NAME/KEY: misc\_binding  
LOCATION: 481..500  
OTHER INFORMATION: 10-507-364.misl, potential  
NAME/KEY: misc\_binding  
LOCATION: 502..521  
OTHER INFORMATION: 10-507-364.mis2, potential complement  
NAME/KEY: primer\_bind

LOCATION: 138..156  
OTHER INFORMATION: upstream amplification primer  
NAME/KEY: primer\_bind  
LOCATION: 545..564  
OTHER INFORMATION: downstream amplification primer, complement  
NAME/KEY: misc\_binding  
LOCATION: 489..513  
OTHER INFORMATION: 10-507-364 potential probe  
US-09-641-638-634

Query Match 76.8%; Score 19.2; DB 4; Length 1000;  
Best Local Similarity 87.5%; Pred. No. 14;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAATGGTGGCGGTGGGGGAGAC 25  
||||| ||||| ||||| ||||| |||||  
Db 10 AAATGGTGGCGGGGGGGGAGAC 33

## RESULT 6

US-09-641-638-651  
Sequence 651, Application US/09641638  
Patent No. 6432648  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Annick  
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
FILE REFERENCE: GENSET.051CP1  
CURRENT APPLICATION NUMBER: US/09/641.638  
CURRENT FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: US 09/502,330  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: US 60/133,200  
PRIOR FILING DATE: 1999-05-07  
PRIOR APPLICATION NUMBER: US 09/275,267  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: US 60/119,917  
PRIOR FILING DATE: 1999-02-12  
NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: Patent.pm  
SEQ ID NO 651  
LENGTH: 20674  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1123..3123  
OTHER INFORMATION: 5' regulatory region  
NAME/KEY: exon  
LOCATION: 3124..3297  
OTHER INFORMATION: exon 1  
NAME/KEY: exon  
LOCATION: 3871..4072  
OTHER INFORMATION: exon 2  
NAME/KEY: exon  
LOCATION: 5552..5633  
OTHER INFORMATION: exon 3  
NAME/KEY: exon  
LOCATION: 5758..5880  
OTHER INFORMATION: exon 4  
NAME/KEY: exon  
LOCATION: 5996..6099  
OTHER INFORMATION: exon 5  
NAME/KEY: exon  
LOCATION: 6349..6509  
OTHER INFORMATION: exon 6  
NAME/KEY: exon  
LOCATION: 7379..7522  
OTHER INFORMATION: exon 7  
NAME/KEY: exon

```
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc.feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4052
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
```

NAME/KEY: allele  
LOCATION: 13535

Query Match 76.8%; Score 19.2; DB 4; Length 20674;  
Best Local Similarity 87.5%; Pred. No. 16;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAATGGTGGCGGTGGGGGAGAC 25  
||||| ||||| ||||| ||||| |||||

Db 13044 AAATGGTGGCGGTGGGGGAGAC.13067

## RESULT 7

US-08-680-395-4/c  
Sequence 4, Application US/08680395

Patent No. 5892010

## GENERAL INFORMATION:

APPLICANT: Gray, Joe W.  
APPLICANT: Collins, Colin  
APPLICANT: Hwang, Soo-in  
APPLICANT: Godfrey, Tony  
APPLICANT: Kowbel, David  
APPLICANT: Rommens, Johanna  
TITLE OF INVENTION: Genes from the 20q13 Amlicon and Their  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

ZIP: 94111-3834

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/680.395  
APPLICATION NUMBER: US/08/680.395

FILING DATE: 15-JUL-1996

CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-0689000US

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2605 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

## FEATURE:

NAME/KEY: 1..2605

LOCATION: /note= "cDNA clone cc43 of 4 kb

OTHER INFORMATION: transcript"

US-08-680-395-4

Query Match 74.4%; Score 18.6; DB 2; Length 2605;  
Best Local Similarity 84.0%; Pred. No. 25;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TAAATGGTGGCGGTGGGGGAGAC 25

||||| ||||| ||||| ||||| |||||

Db 1015 TAAAGGGTGGGAGGAGGAGAC 991

## RESULT 8

US-07-674-287B-1/c  
Sequence 1, Application US/07674287B

Patent No. 5414076

## GENERAL INFORMATION:

APPLICANT: Bryan Mark O'Hara  
TITLE OF INVENTION: Gibbon Ape Leukemia  
TITLE OF INVENTION: Virus Receptor  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dr. Karen A. Lowney

ADDRESSEE: American Cyanamid Company

STREET: 1937 West Main Street

STREET: P.O. Box 60

CITY: Stamford

STATE: CT

COUNTRY: USA

ZIP: 06904-0060

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC AT

OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII converted from IBM DW4

CURRENT APPLICATION DATA:

TITLE OF INVENTION: US/07/674,287B

APPLICATION NUMBER: 19910325

FILING DATE: 19910325

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

## ATTORNEY/AGENT INFORMATION:

NAME: Lowney, Karen A., Dr.

REGISTRATION NUMBER: 31,274

REFERENCE/DOCKET NUMBER: 31,104-01

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 203 321 2361

TELEFAX: 203 321 2971

TELEX: 710 474 4059

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3211 Base Pairs

TYPE: NUCLEOTIDE SEQUENCE

STRANDEDNESS: Single

TOPOLOGY: Linear

MOLECULE TYPE: DNA

US-07-674-287B-1

Query Match 70.4%; Score 17.6; DB 1; Length 3211;

Best Local Similarity 83.3%; Pred. No. 64;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATGGTGGCGGTGGGGGAGAC 25

||||| ||||| ||||| ||||| |||||

Db 99 AAAAGGAGGCGGAGCGGAGAC 76

## RESULT 9

US-08-436-900A-1/c

Sequence 1, Application US/08436900A

Patent No. 5874264

## GENERAL INFORMATION:

APPLICANT: O'Hara, Bryan M.

TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Home Products

STREET: One Campus Drive

CITY: Parsippany

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07054

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/436,900A  
FILING DATE: 08-MAY-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Barnhard, Elizabeth M.  
REGISTRATION NUMBER: 31,088  
REFERENCE/DOCKET NUMBER: 31,104-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-683-2158  
TELEFAX: 201-683-4117  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3211 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-436-900A-1

Query Match 70.4%; Score 17.6; DB 2; Length 3211;  
Best Local Similarity 83.3%; Pred. No. 64;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATGGTGGCGGTGGGGAGAC 25  
||| ||| ||||| | |||||  
Db 99 AAAGGGAGCGGCGGAGAC 76

RESULT 10  
US-08-171-389-505/c  
Sequence 505, Application US/08171389  
Patent No. 5578444  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 641  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/171,389  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:

NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0175/G19P3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 505:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene  
INDIVIDUAL ISOLATE: (start site 106547)  
US-08-171-389-505  
Query Match 68.0%; Score 17; DB 1; Length 50;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGGCGGTGGGGAGAC 24  
||||| ||||| |||||  
Db 50 GTGGCGGTGGGGAGAC 34

RESULT 11  
US-08-123-936-505/c  
Sequence 505, Application US/08123936  
Patent No. 5726014  
GENERAL INFORMATION:  
APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
TITLE OF INVENTION: Screening Assay for the Detection of  
TITLE OF INVENTION: DNA-Binding Molecules  
NUMBER OF SEQUENCES: 640  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,936  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Fabian, Gary R.  
REGISTRATION NUMBER: 33,875  
REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 505:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid

;  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
;  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene  
; INDIVIDUAL ISOLATE: (start site 106547)  
US-08-123-936-505

Query Match 68.0%; Score 17; DB 1; Length 50;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGGCGGTGGGGGAGA 24  
|||||

DB 50 GTGGCGGTGGGGGAGA 34

RESULT 12  
US-08-475-228A-505/c  
; Sequence 505, Application US/08475228A  
; Patent No. 5869241  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,228A  
; FILING DATE: 06-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stratford, Carol A.  
; REGISTRATION NUMBER: 34,444  
; REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P302  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 505:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO

;  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene  
; INDIVIDUAL ISOLATE: (start site 106547)  
US-08-475-228A-505

Query Match 68.0%; Score 17; DB 2; Length 50;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGGCGGTGGGGGAGA 24  
|||||

DB 50 GTGGCGGTGGGGGAGA 34

RESULT 13  
US-08-482-080A-505/c  
; Sequence 505, Application US/08482080A  
; Patent No. 6010849  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,080A  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/171,389  
; FILING DATE: 20-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brady, John F.  
; REGISTRATION NUMBER: 39,118  
; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 324-0880  
; TELEFAX: (650) 324-0960  
; INFORMATION FOR SEQ ID NO: 505:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:

; ; INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene  
; ; INDIVIDUAL ISOLATE: (start site 106547)  
US-08-482-080A-505

Query Match 68.0%; Score 17; DB 3; Length 50;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGGCGGTGGGGGAGA 24  
|||||  
Db 50 GTGGCGGTGGGGGAGA 34

## RESULT 14

US-09-354-947-505/c  
; Sequence 505, Application US/09354947  
; Patent No. 6384208  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Cynthia A.  
; APPLICANT: Cantor, Charles R.  
; APPLICANT: Andrews, Beth M.  
; APPLICANT: Turin, Lisa M.  
; APPLICANT: Fry, Kirk E.  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 664  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/354,947  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/482,080  
; FILING DATE: 07-JUN-1995  
; APPLICATION NUMBER: US 08/171,389  
; FILING DATE: 20-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/723,618  
; FILING DATE: 27-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/081,070  
; FILING DATE: 22-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brady, John F.  
; REGISTRATION NUMBER: 39,118  
; REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 324-0880  
; TELEFAX: (650) 324-0960  
; INFORMATION FOR SEQ ID NO: 505:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO

; ; ORIGINAL SOURCE:  
; ; INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene  
; ; INDIVIDUAL ISOLATE: (start site 106547)  
US-09-354-947-505

Query Match 68.0%; Score 17; DB 4; Length 50;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGGCGGTGGGGGAGA 24  
|||||  
Db 50 GTGGCGGTGGGGGAGA 34

## RESULT 15

PCT-US93-12388-505/c  
; Sequence 505, Application PC/TUS9312388  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Sequence-Directed DNA Binding  
; TITLE OF INVENTION: Molecules, Compositions and Methods  
; NUMBER OF SEQUENCES: 641  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genelabs Technologies, Inc.  
; STREET: 505 Penobscot Drive  
; CITY: Redwood City  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94063  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US93/12388  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/123,936  
; FILING DATE: 17-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/996,783  
; FILING DATE: 23-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 505:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 50 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ; INDIVIDUAL ISOLATE: Herpes simplex virus-1 b' 42K gene  
; ; INDIVIDUAL ISOLATE: (start site 106547)  
PCT-US93-12388-505

Query Match 68.0%; Score 17; DB 5; Length 50;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GTGGCGGTGGGGGAGA 24  
|||||  
Db 50 GTGGCGGTGGGGGAGA 34



Search completed: January 19, 2003, 01:35:38  
Job time : 52 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:37:54 ; Search time 46 Seconds  
(without alignments)  
166.672 Million cell updates/sec

Title: US-09-674-593-9

Perfect score: 25

Sequence: 1 taggctgttggaaagggtagcaca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 9  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

100% processing. Minimum Match 0%  
Maximum Match 100%

## Listing first 45 summaries

Database : Issued patents NA: \*

```
1: /cgn2_6/ptodata/2/1na/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq
6: /cgn2_6/ptodata/2/1na/pctfiles1.seq
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	18.2	72.8	2055	4	US-09-197-218-1		Sequence 1, Appl
2	17.8	71.2	1491	6	5486473-3		Patent No. 5486473
C 3	17	68.0	1309	3	US-08-362-495-4		Sequence 4, Appl
C 4	17	66.4	1309	4	US-09-408-508-4		Sequence 4, Appl
C 5	16.6	66.4	9711	4	US-08-961-527-1		Sequence 167, App
C 6	16.2	64.8	849	4	US-08-998-416-552		Sequence 552, App
7	16.2	64.8	1331	3	US-08-985-908-23		Sequence 23, Appl
8	16.2	64.8	9096	4	US-09-147-119-5		Sequence 5, Appl
9	16.2	64.8	15936	4	US-09-147-119-1		Sequence 1, Appl
C 10	16	64.0	540	4	US-08-961-527-288		Sequence 288, App
C 11	16	64.0	4129	2	US-08-370-319C-12		Sequence 12, Appl
C 12	16	64.0	4129	4	US-09-224-834-12		Sequence 12, Appl
13	16	64.0	4503	2	US-08-770-301A-2		Sequence 2, Appl
14	16	64.0	4503	3	US-09-175-581-2		Sequence 2, Appl
C 15	16	64.0	16389	4	US-09-741-154-3		Sequence 3, Appl
C 16	16	64.0	26385	4	US-08-961-527-3		Sequence 3, Appl
C 17	15.8	63.2	833	2	US-08-343-443B-97		Sequence 97, Appl
18	15.8	63.2	1488	2	US-08-687-559-5		Sequence 5, Appl
19	15.8	63.2	2233	4	US-09-484-970B-141		Sequence 141, App
20	15.8	63.2	9495	1	US-08-271-829-1		Sequence 1, Appl
21	15.8	63.2	9495	5	PCR-US93-01544-1		Sequence 1, Appl
22	15.6	62.4	539	3	US-08-828-741B-12		Sequence 12, Appl
23	15.6	62.4	539	4	US-09-160-567-12		Sequence 12, Appl
24	15.6	62.4	548	3	US-08-828-741B-1		Sequence 1, Appl
25	15.6	62.4	548	4	US-09-160-567-1		Sequence 1, Appl
26	15.6	62.4	599	3	US-08-828-741B-7		Sequence 7, Appl
27	15.6	62.4	599	4	US-09-160-567-7		Sequence 7, Appl

28	15.6	62.4	1031	3	US-08-828-741B-5	Sequence 5, Appl 1
29	15.6	62.4	1031	4	US-09-160-567-5	Sequence 5, Appl 1
30	15.6	62.4	1490	3	US-08-828-741B-3	Sequence 3, Appl 1
31	15.6	62.4	1490	4	US-09-160-567-3	Sequence 3, Appl 1
32	15.6	62.4	2042	4	US-08-869-423-4	Sequence 4, Appl 1
33	15.4	61.6	632	4	US-09-328-111-124	Sequence 124, App
34	15.4	61.6	1043	4	US-09-165-868-4	Sequence 4, Appl 1
C 35	15.4	61.6	1160	6	5169941-2	Patent No. 5169941
C 36	15.4	61.6	1512	3	US-08-945-994-4	Sequence 4, Appl 1
C 37	15.4	61.6	1788	1	US-08-325-989-1	Sequence 1, Appl 1
C 38	15.4	61.6	1788	1	US-08-370-923-1	Sequence 1, Appl 1
C 39	15.4	61.6	1788	1	US-08-380-014-1	Sequence 1, Appl 1
C 40	15.4	61.6	1788	3	US-09-079-785-1	Sequence 1, Appl 1
41	15.4	61.6	1890	3	US-09-289-466-2	Sequence 2, Appl 1
42	15.4	61.6	1891	3	US-09-289-466-1	Sequence 1, Appl 1
43	15.4	61.6	1929	2	US-09-016-000-10	Sequence 10, Appl 1
C 44	15.4	61.6	2932	2	US-08-481-337A-5	Sequence 5, Appl 1
C 45	15.4	61.6	2932	4	US-09-382-256-5	Sequence 5, Appl 1

## ALIGNMENTS

```

RESULT 1
US-09-197-218-1
; Sequence 1, Application US/09197218A
; Patent No. 6258788
; GENERAL INFORMATION:
; APPLICANT: Schmaljohn, Connie S.
; TITLE OF INVENTION: DNA Vaccines Against Tick-borne Flaviviruses
; FILE REFERENCE: 003/0100/SAP RLID 96-10
; CURRENT APPLICATION NUMBER: US/09/197,218A
; CURRENT FILING DATE: 1998-11-20
; EARLIER APPLICATION NUMBER: US 60/065,750
; EARLIER FILING DATE: 1997-11-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 1
; LENGTH: 2055
; TYPE: DNA
; ORGANISM: Central European Encephalitis virus
; FEATURE:
; OTHER INFORMATION: genomic
US-09-197-218-1

```

```

Query Match      72.8%; Score 18.2; DB 4; Length 2055;
Best Local Similarity 87.0%; Pred. No. 6.1;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGCTGCTTTGGAAAGGGTAGCA 23
      | | | | | | | | | | | | | |
DB      864 TGGACTTTTGGAAAGGGTAGCA 886

RESULT 2
5486473-3
; Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI, YOSHIDA, IWAO, TAKAGI, MITSUO;
; MANABE, SADAO; FUKAI, KONOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
; SEQ ID NO: 3
; SEQ LENGTH: 1491

```



Query Match 66.4%; Score 16.6; DB 4; Length 9711;  
Best Local Similarity 82.6%; Pred. No. 53;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TAGGCTGTTTGGAAAGGTAGCA 23  
I IIII IIII II IIIIIII  
Db 1972 TTGGCTTTTGGGAATGTCAGCA 1950

## RESULT 6

US-08-998-416-552  
; Sequence 552, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reibischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CCCL976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 552:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 849 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1384UP  
; US-08-998-416-552

Query Match 64.8%; Score 16.2; DB 4; Length 849;  
Best Local Similarity 85.7%; Pred. No. 48;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 CTGTTTGGAAAGGTAGCACA 25  
IIIIII III IIII IIIII  
Db 245 CTGTTTGAATGGGTGGCACA 265

## RESULT 7

US-08-985-908-23  
; Sequence 23, Application US/08985908

; Patent No. 6004773  
; GENERAL INFORMATION:  
; APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI  
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.  
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
; CITY: ARLINGTON  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/985,908  
; FILING DATE: 05-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-325659  
; FILING DATE: 05-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NORMAN F. OBLON  
; REGISTRATION NUMBER: 24,618  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1331 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: JM109  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 10..1197  
; US-08-985-908-23

Query Match 64.8%; Score 16.2; DB 3; Length 1331;  
Best Local Similarity 85.7%; Pred. No. 54;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GCTGTTTGGAAAGGTAGCAC 24  
IIIIIIII II IIIII  
Db 252 GCTGTTGGTAAAGGTAGCGC 272

## RESULT 8

US-09-147-119-5  
; Sequence 5, Application US/09147119  
; Patent No. 6338844  
; GENERAL INFORMATION:  
; APPLICANT: KURTH, Ruth  
; APPLICANT: BAIER, Michael  
; APPLICANT: BANNERT, No. 6338844bert  
; APPLICANT: METZNER, Karin  
; APPLICANT: WERNER, Albrecht  
; TITLE OF INVENTION: GENOMIC NUCLEIC ACIDS, cDNA AND mRNA WHICH  
; TITLE OF INVENTION: CODE FOR POLYPEPTIDES WITH IL-16 ACTIVITY, PROCESS FOR THE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP  
; STREET: 655 Fifteenth St., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA

ZIP: 20005-5701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/147,119  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP97/01753  
FILING DATE: 10-APR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 196 14 099.4  
FILING DATE: 10-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Wong, King L.  
REGISTRATION NUMBER: 37,500  
REFERENCE/DOCKET NUMBER: P1614-8068  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-5000  
TELEFAX: (202) 638-4810  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9096 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: exon  
LOCATION: 1..338  
FEATURE:  
NAME/KEY: intron  
LOCATION: 339..663  
FEATURE:  
NAME/KEY: exon  
LOCATION: 664..832  
FEATURE:  
NAME/KEY: intron  
LOCATION: 833..2870  
FEATURE:  
NAME/KEY: exon  
LOCATION: 2871..2972  
FEATURE:  
NAME/KEY: intron  
LOCATION: 2973..5224  
FEATURE:  
NAME/KEY: exon  
LOCATION: 5225..5483  
FEATURE:  
NAME/KEY: intron  
LOCATION: 5484..5737  
FEATURE:  
NAME/KEY: exon  
LOCATION: 5738..5863  
FEATURE:  
NAME/KEY: intron  
LOCATION: 5864..7926  
FEATURE:  
NAME/KEY: exon  
LOCATION: 7927..9096  
FEATURE:  
NAME/KEY: -  
LOCATION: 356  
OTHER INFORMATION: /product= "N means between 1 -  
OTHER INFORMATION: about 6 bp"  
US-09-147-119-5

Query Match 64.8%; Score 16.2; DB 4; Length 9096;  
Best Local Similarity 85.7%; Pred. No. 82;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GGCTGTTTGGAAAGGGTAGCA 23  
||||| ||||| ||||| |||||  
DB 3132 GGCTGCTGGAGAGGGTAGCA 3152  
RESULT 9  
US-09-147-119-1  
; Sequence 1, Application US/09147119  
; Patent No. 6338844  
; GENERAL INFORMATION:  
; APPLICANT: KURTH, Ruth  
; APPLICANT: BAIER, Michael  
; APPLICANT: BANNERT, No. 6338844bert  
; APPLICANT: METZNER, Karin  
; APPLICANT: WERNER, Albrecht  
; TITLE OF INVENTION: GENOMIC NUCLEIC ACIDS, CDNA AND mRNA WHICH  
; TITLE OF INVENTION: CODE FOR POLYPEPTIDES WITH IL-16 ACTIVITY, PROCESS FOR THE  
; TITLE OF INVENTION: PRODUCTION THEREOF AND THEIR USE  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaïdo, Marmelstein, Murray & Oram LLP  
; STREET: 655 Fifteenth St., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005-5701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/147,119  
; FILING DATE: HEREWITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP97/01753  
; FILING DATE: 10-APR-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 196 14 099.4  
; FILING DATE: 10-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wong, King L.  
; REGISTRATION NUMBER: 37,500  
; REFERENCE/DOCKET NUMBER: P1614-8068  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-5000  
; TELEFAX: (202) 638-4810  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15936 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 3100..3238  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 5540..6635  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 7504..7672  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 9711..9812  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: 12065..12323  
; FEATURE:  
; NAME/KEY: exon

```
; LOCATION: 12578..12703
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14767..15936
US-09-147-119-1
Query Match 64.8%; Score 16.2; DB 4; Length 15936;
Best Local Similarity 85.7%; Pred. No. 93;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GCCTGTTTGGAAAGGTCACCA 23
Db 9972 GCCTGCTGGAGAGGTCACCA 9992

RESULT 10
US-08-961-527-288/c
; Sequence 288, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-288
Query Match 64.0%; Score 16; DB 4; Length 540;
Best Local Similarity 79.2%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGCGTGTTCGAAAGGTCAGCACA 25
Db 161 AGCGTGTTCGAAAGGTCGCAAA 138

RESULT 11
US-08-370-319C-12/c
; Sequence 12, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE
```

```
; APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE
; TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,319C
; FILING DATE: 10-JANUARY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/272,351
; FILING DATE: 8-JULY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/032,978
; FILING DATE: 18-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5856091man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: The sequence is preceded by an
; OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
; OTHER INFORMATION: kilobases
US-08-370-319C-12
Query Match 64.0%; Score 16; DB 2; Length 4129;
Best Local Similarity 79.2%; Pred. No. 86;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGCGTGTTCGAAAGGTCAGCACA 25
Db 732 AGTCAGTTTGGAAAGTATACCACA 709

RESULT 12
US-09-224-834-12/c
; Sequence 12, Application US/09224834
; Patent No. 6201111
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W lfel, Thomas; Coullie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE
; TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
```

```

RESULT 13
US-08-770-301A-2
: Sequence 2, Application US/08770301A
: Patent No. 5948637
: GENERAL INFORMATION:
: APPLICANT: IKEDA, JUN
: APPLICANT: KANEDA, SUMIKO
: APPLICANT: YANAGI, HIDEKI
: APPLICANT: MATSUMOTO, MASAYASU
: APPLICANT: YURA, TAKASHI
: TITLE OF INVENTION: NOVEL STRESS PROTEINS
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
: STREET: PO BOX 747
: CITY: FALLS CHURCH
: STATE: VA
: COUNTRY: USA
: ZIP: 22040-0747
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/770,301A
: FILING DATE: 20-DEC-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:

```

```

; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1422-287
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-205-8000
; TELEFAX: (703)-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; IDENTIFICATION METHOD: E
; US-08-770-301A-2

Query Match 64.0%; Score 16; DB 2; Length 4503;
Best Local Similarity 79.2%; Pred. No. 88;
Matches 19; Conservative 0; Mismatches 5; Indels

QY 2 AGCGTGTGGAAAGGGTAGCACA 25
|| ||||| |||| ||||
Db 1849 AGCGTGTGGAGCGGTACCACA 1872

RESULT 14
US-09-175-581-2
; Sequence 2, Application US/09175581
; Patent No. 6034232
; GENERAL INFORMATION:
; APPLICANT: IKEDA, JUN
; APPLICANT: KANEDA, SUMIKO
; APPLICANT: YANAGI, HIDEKI
; APPLICANT: MATSUMOTO, MASAYASU
; APPLICANT: YURA, TAKASHI
; TITLE OF INVENTION: NOVEL STRESS PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/175,581
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/770,301
; FILING DATE: 20-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1422-287
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-205-8000
; TELEFAX: (703)-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4503 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:

```



```
; NAME/KEY: CDS
; IDENTIFICATION METHOD: E
US-09-175-581-2

Query Match      64.08; Score 16; DB 3; Length 4503;
Best Local Similarity 79.2%; Pred. No. 88;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy  2 AGGCTGTTTGGAAAGGGTAGCACCA 25
    ||||||||| ||| |||
Db  1849 AGCCTGTTTGGAGGCGGTACCACA 1872

RESULT 15
US-09-741-154-3
; Sequence 3, Application US/09741154
; Patent No. 6437110
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01061
; CURRENT APPLICATION NUMBER: US/09/741.154
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 16389
; TYPE: DNA
; ORGANISM: Human
US-09-741-154-3

Query Match      64.08; Score 16; DB 4; Length 16389;
Best Local Similarity 79.2%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy  2 AGGCTGTTTGGAAAGGGTAGCACCA 25
    ||||||| ||| ||| ||| |||
Db  16205 AGGCTGCGAGAGGGGTACCACA 16228

Search completed: January 19, 2003, 01:35:45
Job time : 53 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:44:49 ; Search time 57.5 seconds  
(without alignments)  
193.856 Million cell updates/sec

Title: US-09-674-593-8

Perfect score: 25

Sequence: 1 taaatgggtggcggtggggagac 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 393868 seqs, 222934149 residues

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.8	75.2	639	US-09-938-842A-1976	Sequence 1976, Ap
C 2	18.2	72.8	853	US-09-764-864-72	Sequence 72, Appl
C 3	18.2	72.8	23433	US-09-927-091-7	Sequence 7, Appl
C 4	18.2	72.8	30825	US-09-927-091-5	Sequence 5, Appl
C 5	17.6	70.4	273	US-09-923-876-5830	Sequence 5830, Ap
6	17.6	70.4	276	US-09-728-446-670	Sequence 670, App
7	17.6	70.4	501	US-09-878-574-4421	Sequence 4421, Ap
C 8	17.6	70.4	3220	US-09-954-531-137	Sequence 137, App
C 9	17.6	70.4	3220	US-09-954-531-356	Sequence 356, App
C 10	17.6	70.4	3290	US-09-981-353-172	Sequence 172, App
C 11	17.2	68.8	345	US-09-974-300-211	Sequence 211, App
12	17.2	68.8	170834	US-09-835-232-7	Sequence 7, Appl
C 13	17	68.0	268	US-10-040-739-312	Sequence 312, App
C 14	17	68.0	390	US-09-954-456-1960	Sequence 1960, Ap
15	17	68.0	390	US-09-880-107-2872	Sequence 2872, Ap
16	17	68.0	1503841	US-09-946-807-1	Sequence 1, Appl
17	17	68.0	1503841	US-09-795-668-1	Sequence 1, Appl
18	17	68.0	1503841	US-09-795-686-1	Sequence 1, Appl
C 19	16.8	67.2	305	US-09-864-761-19262	Sequence 19262, A

C 20	16.8	67.2	463	10	US-09-563-817-536	Sequence 536, App
C 21	16.8	67.2	496	10	US-09-864-761-2534	Sequence 2534, Ap
C 22	16.8	67.2	198285	10	US-09-880-107-3814	Sequence 3814, Ap
C 23	16.6	66.4	404	10	US-09-960-352-9889	Sequence 9889, Ap
C 24	16.6	66.4	405	9	US-09-768-827-20	Sequence 20, Appl
25	16.6	66.4	427	10	US-09-960-352-5580	Sequence 5580, Ap
26	16.6	66.4	1095	10	US-09-728-628-2	Sequence 2, Appl
27	16.6	66.4	1379	10	US-09-822-849A-99	Sequence 99, Appl
C 28	16.6	66.4	1827	10	US-09-925-297-180	Sequence 180, App
C 29	16.6	66.4	2214	10	US-09-880-107-3285	Sequence 3285, Ap
30	16.6	66.4	2523	10	US-09-880-107-2167	Sequence 2167, Ap
31	16.6	66.4	2595	10	US-09-919-497-12	Sequence 12, Appl
32	16.6	66.4	2901	12	US-10-044-090-528	Sequence 528, App
33	16.6	66.4	4105	10	US-09-931-157-1	Sequence 1, Appl
34	16.6	66.4	7739	10	US-09-764-877-3189	Sequence 3189, Ap
35	16.6	66.4	37950	9	US-09-853-526-183	Sequence 183, App
36	16.6	66.4	37950	10	US-09-901-484A-183	Sequence 183, App
37	16.6	66.4	57130	10	US-09-835-081-3	Sequence 3, Appl
C 38	16.4	65.6	1364	10	US-09-822-849A-19	Sequence 19, Appl
C 39	16.4	65.6	1524	10	US-09-878-262B-4	Sequence 4, Appl
C 40	16.4	65.6	2466	12	US-10-044-090-251	Sequence 251, App
C 41	16.4	65.6	2469	10	US-09-878-262B-6	Sequence 6, Appl
C 42	16.4	65.6	65359	10	US-09-804-472-3	Sequence 3, Appl
C 43	16.2	64.8	208	9	US-09-933-797-680	Sequence 680, App
C 44	16.2	64.8	286	10	US-09-923-876-3994	Sequence 3994, Ap
C 45	16.2	64.8	455	10	US-09-416-384A-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1

US-09-938-842A-1976/C

; Sequence 1976, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

; FILE REFERENCE: SAME, AND METHODS OF USE

; CURRENT APPLICATION NUMBER: US/09/938,842A

; PRIOR FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227,866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264,647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300,111

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 1976

; LENGTH: 639

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-1976

Query Match 75.2%; Score 18.8; DB 9; Length 639;

Best Local Similarity 90.9%; Pred. No. 31;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATGGTGGCGGTGGGGGAGA 24

||||| ||||| ||||| |||||

Db 634 AATGTGTGGAGGTGGGGGAGA 613

RESULT 2

US-09-764-864-72/c

; Sequence 72, Application US/09764864

; Patent No. US20020132753A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT23  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 72  
; LENGTH: 853  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-864-72

Query Match 72.8%; Score 18.2; DB 10; Length 853;  
Best Local Similarity 87.0%; Pred. No. 55;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATGGGTGGCGGTGGGGGAGA 24  
||| ||||| ||||| ||  
Db 608 AAAAGGTGGGGGTGGGGGGA 586

## RESULT 3

US-09-927-091-7/c  
; Sequence 7, Application US/09927091  
; Patent No. US20020119541A1  
; GENERAL INFORMATION:  
; APPLICANT: KILLARY, ANN  
; APPLICANT: LOTT, STEVE  
; APPLICANT: CHANDLER, DAWN  
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1  
; FILE REFERENCE: UTSC:651US  
; CURRENT APPLICATION NUMBER: US/09/927,091  
; CURRENT FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/227,560  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: 60/225,033  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 23433  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (5071)..(23433)  
; OTHER INFORMATION: n = A or C or G or T/U  
US-09-927-091-7

Query Match 72.8%; Score 18.2; DB 10; Length 23433;  
Best Local Similarity 87.0%; Pred. No. 63;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAG 23  
||| ||||| ||||| ||  
Db 7223 TACAGGGTGGGAGGTGGGGGAG 7201

## RESULT 4

US-09-927-091-5/c  
; Sequence 5, Application US/09927091  
; Patent No. US20020119541A1  
; GENERAL INFORMATION:  
; APPLICANT: KILLARY, ANN  
; APPLICANT: LOTT, STEVE  
; APPLICANT: CHANDLER, DAWN  
; TITLE OF INVENTION: THE TUMOR SUPPRESSOR CAR-1  
; FILE REFERENCE: UTSC:651US  
; CURRENT APPLICATION NUMBER: US/09/927,091  
; CURRENT FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 60/227,560  
; PRIOR FILING DATE: 2000-08-23

; PRIOR APPLICATION NUMBER: 60/225,033  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 30625  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (4754)..(30625)  
; OTHER INFORMATION: n = A or C or G or T/U  
US-09-927-091-5

Query Match 72.8%; Score 18.2; DB 10; Length 30625;  
Best Local Similarity 87.0%; Pred. No. 64;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAG 23  
||| ||||| ||||| ||  
Db 16649 TACAGGGTGGGAGGTGGGGGAG 16627

## RESULT 5

US-09-923-876-5830/c  
; Sequence 5830, Application US/09923876  
; Patent No. US20020013958A1  
; GENERAL INFORMATION:  
; APPLICANT: Laligudi, Raghunath V.  
; APPLICANT: Kamigaki, Laura Y. (Ito)  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING  
; FILE REFERENCE: PL-0012-1 CON  
; CURRENT APPLICATION NUMBER: US/09/923,876  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: 09/298,329  
; PRIOR FILING DATE: 1999-04-21  
; PRIOR APPLICATION NUMBER: 60/085,331  
; PRIOR FILING DATE: 1998-05-05  
; NUMBER OF SEQ ID NOS: 6332  
; SOFTWARE: PERL Program  
; SEQ ID NO 5830  
; LENGTH: 273  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700457813H1  
; NAME/KEY: unsure  
; LOCATION: 10, 44, 196  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-923-876-5830

Query Match 70.4%; Score 17.6; DB 10; Length 273;  
Best Local Similarity 83.3%; Pred. No. 90;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATGGGTGGCGGTGGGGGAGAC 25  
||| | ||| ||||| ||||  
Db 97 AAATTGATGGGTGGTGGGGCAGAC 74

## RESULT 6

US-09-728-446-670  
; Sequence 670, Application US/09728446  
; Patent No. US20020081668A1  
; GENERAL INFORMATION:  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. US20020081668A1 Murine Polynucleotide Sequences  
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby  
; FILE REFERENCE: LEX-0101-USA

;  
; CURRENT APPLICATION NUMBER: US/09/728,446  
; CURRENT FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/168,270  
; PRIOR FILING DATE: 1999-12-01  
; NUMBER OF SEQ ID NOS: 1461  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 670  
; LENGTH: 276  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(276)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-728-446-670

Query Match 70.4%; Score 17.6; DB 10; Length 276;  
Best Local Similarity 83.3%; Pred. No. 90;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAGA 24  
||||||| ||||| ||||| ||  
Db 227 TAAATGGGTGGCGGGGGGGGA 250

RESULT 7

US-09-878-574-4421  
; Sequence 4421, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 4421  
; LENGTH: 501  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(501)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB3028-022-Q1-B1-C2  
US-09-878-574-4421

Query Match 70.4%; Score 17.6; DB 10; Length 501;  
Best Local Similarity 83.3%; Pred. No. 92;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAGA 24  
||| ||||| || ||||| ||  
Db 363 TGAAGGGGTGGGGGAGGGGGGA 386

RESULT 8

US-09-954-531-137/c  
; Sequence 137, Application US/09954531  
; Patent No. US20020165180A1  
; GENERAL INFORMATION:  
; APPLICANT: Weaver, zoe  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
; FILE REFERENCE: 689290-77  
; CURRENT APPLICATION NUMBER: US/09/954,531  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US/60/233,133

;  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,009  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,034  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,509  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US/60/234,567  
; PRIOR FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 1392  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 137  
; LENGTH: 3220  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-531-137

Query Match 70.4%; Score 17.6; DB 9; Length 3220;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATGGGTGGCGGTGGGGGAGAC 25  
||| ||| ||||| ||||| ||  
Db 99 AAAAGGGAGGCGGAGCGGGAGAC 76

RESULT 9

US-09-954-531-356/c  
; Sequence 356, Application US/09954531  
; Patent No. US20020165180A1  
; GENERAL INFORMATION:  
; APPLICANT: Weaver, zoe  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; TITLE OF INVENTION: Gene Sets  
; FILE REFERENCE: 689290-77  
; CURRENT APPLICATION NUMBER: US/09/954,531  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US/60/233,133  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,009  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,034  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,509  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US/60/234,567  
; PRIOR FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 1392  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 356  
; LENGTH: 3220  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-531-356

Query Match 70.4%; Score 17.6; DB 9; Length 3220;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATGGGTGGCGGTGGGGGAGAC 25  
||| ||| ||||| ||||| ||  
Db 99 AAAAGGGAGGCGGAGCGGGAGAC 76

RESULT 10

US-09-981-353-172/c  
; Sequence 172, Application US/09981353  
; Patent No. US20020160382A1  
; GENERAL INFORMATION:  
; APPLICANT: Lasek, Amy W.  
; APPLICANT: Jones, David A.  
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
; FILE REFERENCE: PA-0038 US

```

; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 172
; LENGTH: 3290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020
US-09-981-353-172

```

Query Match 70.4%; Score 17.6; DB 9; Length 3290;  
Best Local Similarity 83.3%; Pred. No. 1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0

QY      2    AAATGGGTGGCGGTGGGGGAGAC    25  
           ||| ||| ||||| | |||||  
Db    172   AAAAGGGAGGGCGGAGCGGGAGAC    149

```

RESULT 11
US-09-974-300-211/c
; Sequence 211, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 345
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-211

```

```
Query Match 68.8%; Score 17.2; DB 10; Length 345;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0
```

QY 3 AATGGGTGGCGGTGGGGGAGA 24  
||||| ||| || |||||  
Db 270 AATGGGGGGGTGGGGGGGAGA 249

```

RESULT 12
US-09-835-232-7
; Sequence 7, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
; APPLICANT: Leder, Philip
; APPLICANT: Leder, Philip
; TITLE OF INVENTION: FORMIN-2, Benjamin
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 00383/052002
; CURRENT APPLICATION NUMBER: US/09/835, 232
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,811
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 170834

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(170834)
; OTHER INFORMATION: n= A,T,C, or G
US-09-835-232-7

```

```
Query Match      68.8%; Score 17.2; DB 10; Length 170834;
Best Local Similarity 86.4%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0;
```

QY 1 TAAATGGGTGGCGGTGGGGA 22  
|||||  
Db 111814 TAAATGGGTGGCGAGGTGGA 111835

```

1  RESULT 13
2  US-10-040-7339-312/c
3  ; Sequence 312, Application US/10040739
4  ; Patent No. US20020173635A1
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Jacobs, Kenneth
7  ; McCoy, John
8  ; LaVallie, Edward
9  ; Racie, Lisa
10 ; Merberg, David
11 ; Treacy, Maurice
12 ; Spaulding, Vikki
13 ;
14 ; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
15 ;
16 ; NUMBER OF SEQUENCES: 1519
17 ; CORRESPONDENCE ADDRESS:
18 ; ADDRESSEE: Genetics Institute, Inc.
19 ; STREET: 87 CambridgePark Drive
20 ; CITY: Cambridge
21 ; STATE: Massachusetts
22 ; COUNTRY: U.S.A
23 ; ZIP: 02140
24 ;
25 ; COMPUTER READABLE FORM:
26 ; MEDIUM TYPE: Floppy Disk
27 ; COMPUTER: IBM PC Compatible
28 ; OPERATING SYSTEM: PC-DOS/MS-DOS
29 ; SOFTWARE: Patentin Release #1.0, Version #1.30
30 ;
31 ; CURRENT APPLICATION DATA:
32 ; APPLICATION NUMBER: US/10/040,739
33 ; FILING DATE: 07-Jan-2002
34 ; CLASSIFICATION: <Unknown>
35 ;
36 ; PRIOR APPLICATION DATA:
37 ; APPLICATION NUMBER: 09/036,520
38 ; FILING DATE: 03-JUN-1998
39 ; ATTORNEY/AGENT INFORMATION:
40 ; NAME: Brown, Scott A.
41 ; REGISTRATION NUMBER: 32,724
42 ; TELECOMMUNICATION INFORMATION:
43 ; TELEPHONE: (617) 498-8224
44 ; TELEFAX: (617) 876-5851
45 ;
46 ; INFORMATION FOR SEQ ID NO: 312:
47 ; SEQUENCE CHARACTERISTICS:
48 ; LENGTH: 268 base pairs
49 ; TYPE: nucleic acid
50 ; STRANDEDNESS: double
51 ; TOPOLOGY: linear
52 ;
53 ; MOLECULE TYPE: cdna
54 ; SEQUENCE DESCRIPTION: SEQ ID NO: 312:
55 US-10-040-7339-312

```

Query Match 68.0%; Score 17; DB 9; Length 268;  
Best Local Similarity 80.0%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels

**QY**    1 TAAATGGTGGCGGTGGGGGAGAC 25  
         | | | | | | | | | | | | | |  
**pB**    139 TGAATGGGTGTGCCGTGGGGGGCAC 115



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:34:34 ; Search time 1569 Seconds  
(without alignments)  
258.054 Million cell updates/sec

Title: US-09-674-593-8  
Perfect score: 25  
Sequence: 1 taaatgggtggcggtggggagac 25  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	25	100.0	486	14	BM833169 K-EST0107
C 2	25	100.0	845	14	BQ948660 AGENCOURT
C 3	25	100.0	898	14	BQ650550 AGENCOURT
C 4	25	100.0	963	14	BQ650598 AGENCOURT
C 5	25	100.0	1045	12	BG749180 602708139
C 6	24	96.0	879	14	BQ953400 AGENCOURT

C 7	20.4	81.6	302	14	BQ091950
C 8	20.2	80.8	910	10	BE283772
C 9	19.8	79.2	253	10	BB277121
C 10	19.8	79.2	549	12	BG860434
C 11	19.8	79.2	681	12	BG501674
C 12	19.8	79.2	690	17	BH058191
C 13	19.8	79.2	738	17	AQ329897
C 14	19.8	79.2	896	13	BM562440
C 15	19.4	77.6	926	17	CNS0397K
C 16	19.2	76.8	902	14	BM882409
C 17	19.2	76.8	558	17	BH477324
C 18	19.2	76.8	624	10	AW957063
C 19	19.2	76.8	719	17	BH661879
C 20	19.2	76.8	893	12	BE973172
C 21	19.2	76.8	895	17	CNS02P02
C 22	19.2	76.8	966	14	BQ219869
C 23	19.2	76.8	1012	17	CNS05T90
C 24	19	76.0	369	13	B1070676
C 25	19	76.0	434	13	B1069548
C 26	19	76.0	523	13	B1070677
C 27	19	76.0	902	17	AG184729
C 28	18.8	75.2	318	10	AV552048
C 29	18.8	75.2	323	17	AQ067630
C 30	18.8	75.2	332	10	AV543799
C 31	18.8	75.2	345	10	AV543385
C 32	18.8	75.2	345	17	CNS026L8
C 33	18.8	75.2	364	14	T43297
C 34	18.8	75.2	386	10	AV542795
C 35	18.8	75.2	416	10	AV545806
C 36	18.8	75.2	426	14	T88611
C 37	18.8	75.2	428	10	AV549255
C 38	18.8	75.2	433	10	AV548307
C 39	18.8	75.2	435	10	AV546087
C 40	18.8	75.2	435	10	AV547147
C 41	18.8	75.2	446	17	BH883117
C 42	18.8	75.2	452	10	AV546682
C 43	18.8	75.2	456	10	AV541495
C 44	18.8	75.2	466	10	AV544358
C 45	18.8	75.2	466	10	AV545075

ALIGNMENTS

RESULT 1  
BM833169/C  
LOCUS K-EST0107772 S5SNU484s1 Homo sapiens cdna clone S5SNU484s1-12-C04  
DEFINITION 5', mRNA sequence.  
ACCESSION BM833169  
VERSION BM833169.1 GI:19189578  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 12 row: C column: 04  
High quality sequence stop: 486.  
Location/Qualifiers 1. 486

BM833169 486 bp mRNA linear EST 06-MAR-2002  
K-EST0107772 S5SNU484s1 Homo sapiens cdna clone S5SNU484s1-12-C04  
5', mRNA sequence.



## ORIGIN

Query Match 100.0%; Score 25; DB 14; Length 898;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGGAGAC 25  
|||||  
Db 537 TAAATGGTGGCGGTGGGGGAGAC 513

## RESULT 4

BQ650598/c  
LOCUS BQ650598 963 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8207577 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6283187  
5', mRNA sequence.

ACCESSION BQ650598  
VERSION BQ650598.1 GI:21774770  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov

Tissue Procurement: CGAP (Stanford)

cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: LLCM2478 row: m column: 12  
High quality sequence stop: 618.  
Location/Qualifiers

## FEATURES

source

1..963

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6283187"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."  
BASE COUNT 171 a 350 c 244 g 197 t 1 others  
ORIGIN

Query Match 100.0%; Score 25; DB 14; Length 963;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGGAGAC 25  
|||||  
Db 527 TAAATGGTGGCGGTGGGGGAGAC 503

## RESULT 5

BG749180  
LOCUS BG749180 1045 bp mRNA linear EST 15-MAY-2001  
DEFINITION 602708139F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4844744 5',  
mRNA sequence.  
ACCESSION BG749180

## VERSION

KEYWORDS EST.

## SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

## TITLE

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1682 row: n column: 09

High quality sequence stop: 822.  
Location/Qualifiers

source

1..1045

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4844744"  
/clone\_lib="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library. |"  
BASE COUNT 245 a 322 c 300 g 178 t  
ORIGIN

Query Match 100.0%; Score 25; DB 12; Length 1045;

Best Local Similarity 100.0%; Pred. No. 55;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGGAGAC 25

|||||

Db 3 TAAATGGTGGCGGTGGGGGAGAC 27

## RESULT 6

BQ953400

LOCUS BQ953400

DEFINITION AGENCOURT\_8784199 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:6376296

5', mRNA sequence.

ACCESSION BQ953400

VERSION BQ953400.1 GI:22368878

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

## TITLE

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
Plate: LLCM2558 row: e column: 01  
High quality sequence stop: 690.

## FEATURES

source

Location/Qualifiers  
1. .879  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6376296"  
/clone\_lib="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(C). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library. |"  
1 others  
173 t 1 others  
236 g 173 t 1 others  
275 c 236 g 173 t 1 others

BASE COUNT  
ORIGIN

Query Match 96.0%; Score 24; -DB 14; Length 879;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AATGGGTGGCGTGGGGGAGAC 25

|||||

Db 1 AATGGGTGGCGTGGGGGAGAC 24

## RESULT 7

BQ091950/c

LOCUS Ts764 Thellungiella salsauginea ZAP cDNA library Thellungiella  
salsuginea cDNA, mRNA sequence.  
BQ091950

ACCESSION BQ091950.1 GI:20072593

VERSION EST.

KEYWORDS Thellungiella salsauginea.

SOURCE Thellungiella salsauginea.

ORGANISM Thellungiella salsauginea.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Thellungiella.  
1 (bases 1 to 302)

REFERENCE Wang,Z.L., Li,P.H., Sun,Y.F., Zhang,Q., Zhao,Y.X. and Zhang,H.

Expressed sequence tags from a halophyte Thellungiella salsauginea

cDNA library

Unpublished (2000)

Contact: Hui Zhang

Key Laboratory of Plant Stress Research

The Biology Department of Shandong Normal University

No.88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC

Tel: (86)531-2960864

Fax: (86)531-2966954

Email: zhanghesdnu.edu.cn.

## FEATURES

source

Location/Qualifiers  
1. .302  
/organism="Thellungiella salsauginea"  
/db\_xref="taxon:72664"  
/clone\_lib="Thellungiella salsauginea ZAP cDNA library"  
/dev\_stage="seedling"  
/note="Organ: aerial part tissue; Vector: lambda zap;  
Site\_1: EcoRI; Site\_2: XhoI; total RNA extraction from NaCl(200mM) treated Thellungiella salsauginea by RNAGent kit(Promega); mRNA isolation by MESSAGEMAKER kit(GIBCO BRL); directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis kit(STRATAGEN); the ZAP express library by GigapackIII Gold Cloning kit(STRATAGENE)"  
18 g 68 t  
135 c 18 g 68 t

BASE COUNT  
ORIGIN

Query Match 81.6%; Score 20.4; DB 14; Length 302;

Best Local Similarity 95.5%; Pred. No. 1.8e+03;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AATGGGTGGCGTGGGGGAGA 24

|||||

Db 259 AATGGGTGGAGTGGGGGAGA 238

## RESULT 8

BE283772

LOCUS

DEFINITION 601104026F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3501247 5',  
mRNA sequence.

ACCESSION BE283772

VERSION BE283772.1 GI:9160250

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 910)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM8561 row: c column: 08

High quality sequence start: 41

High quality sequence stop: 105.

## FEATURES

source

Location/Qualifiers  
1. .910

/organism="Mus musculus"

/strain="CZECH II"

/db\_xref="taxon:10090"

/clone="IMAGE:3501247"

/clone\_lib="NCI\_CGAP\_Lu29"

/tissue\_type="spontaneous tumor, metastatic to mammary.

Stem cell origin."

/lab\_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

206 a 206 c 278 g 220 t

BASE COUNT

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 910;

Best Local Similarity 88.0%; Pred. No. 2.2e+03;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGGAGAC 25

|||||

Db 556 TGAATGGTGGCGGTGGGTAGAC 580

## RESULT 9

BB277121/c

LOCUS

DEFINITION BB277121 RIKEN full-length enriched, adult retina Mus musculus CDNA  
clone A30002G18 3', mRNA sequence.

ACCESSION BB277121

VERSION BB277121.1 GI:8974142

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 253)  
 Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Arakawa.T., Carninci  
 P., Endo.T., Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N.,  
 Hironaka.T., Horii.F., Ishii.Y., Ishikawa.J., Ishikawa.T., Itoh.M.,  
 Izawa.M., Kadota.K., Kagawa.I., Kai.C., Kawai.J., Kikuchi.N.,  
 Kiyosawa.H., Kojima.Y., Kondo.S., Koya.S., Kurihara.C., Kusakabe.M.,  
 Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,  
 Ono.T., Owa.C., Saito.H., Sakai.C., Sato.K., Shibata.K., Shibata  
 Y., Shigemoto.Y., Shinagawa.A., Shiraki.T., Sogabe.Y., Sugahara.Y.,  
 Suzuki.H., Suzuki.H., Tagawa.A., Takahashi.F., Tomimaga.N., Toya  
 T., Tsunoda.Y., Watahiki.A., Watanabe.S., Yamamura.T., Yamanaka.I.,  
 Yano.R., Yasunishi.A., Yokota.T., Yoshida.K., Yoshiki.A., Yoshino  
 M., Muramatsu.M. and Hayashizaki.Y.  
 RIKEN Mouse ESTs (Konno.H., et al.)  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resgsc.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>,  
 Carninci.P., Nishiyama.Y., Westover.A., Itoh.M., Nagaoka.S., Sasaki  
 N., Okazaki.Y., Muramatsu.M. and Hayashizaki.Y.  
 Thermotabilization and thermoactivation of thermolabile enzymes by  
 trehalose and its application for the synthesis of full length  
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh.M., Kitsuunai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,  
 Tomaru.Y., Carninci.P., Akiyama.J., Shibata.K., Ozawa.Y., Muramatsu.M., Okazaki  
 Y. and Hayashizaki.Y.  
 Automated filtration-based high-throughput plasmid preparation  
 system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci.P. and Hayashizaki.Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
 19-44 (1999)  
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
 further details.

## FEATURES

source

Location/Qualifiers

1. 253  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A93002G18"  
 /clone\_lib="RIKEN full-length enriched, adult retina"  
 /tissue\_type="retina"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5'  
 GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. cDNA went through one round of normalization  
 to Rot = 20.0 and subtraction to Rot = 459.0. Second  
 strand cDNA was prepared with the primer adapter of  
 sequence [5'GAGAGAGATTCGAGTCTAATTAATTAATCCCCCCCCC  
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
 modified pBluescript KS(+) after bulk excision from Lambda  
 FLC I. -Retina RNA was provided by Stefano Gustinich,  
 Department of Neurobiology, Harvard Medical School, 220  
 Longwood Ave., Boston, MA02115, USA, whose assistance we  
 gratefully acknowledge."  
 66 a 81 c 45 g 61 t

BASE COUNT

ORIGIN

Query Match

79.2%; Score 19.8; DB 10; Length 253;

Best Local Similarity 91.3%; Pred. No. 2.9e+03;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AAATGGTGGCGGTGGGGGAGA 24  
 ||||| ||||| ||||| ||||| |||||  
 DB 120 AAATGGTGGCGGTGGGGGAGA 98  
 ||||| ||||| ||||| ||||| |||||  
 RESULT 10  
 BG860434/c 549 bp mRNA linear EST 29-MAY-2001  
 LOCUS 1024070F01.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II  
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.  
 ACCESSION BG860434  
 VERSION BG860434.1 GI:14241618  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii.  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 549)  
 AUTHORS Grossman.A., Davies.J., Federspiel.N., Harris.E., Lefebvre.P.,  
 McDermott.J.P., Silflow.C., Stern.D. and Surzycki.R.  
 Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
 Unicellular System for Analyzing Gene Function and Regulation in  
 Vascular Plants: project phase 2  
 Unpublished (2000)  
 JOURNAL Contact: Charles Hauser  
 COMMENT DCMB Box 91000  
 Duke University  
 Durham, NC 27708-1000  
 Tel: 919 613 8159  
 Fax: 919 613 8177  
 Email: chauser@duke.edu.  
 FEATURES  
 source  
 1. 549  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap  
 II"  
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
 XhoI; This library, constructed by John Davies and Jeffrey  
 McDermott, combines cDNAs from CC-1690 cells grown to  
 mid-log phase in TAP (acetate-containing) medium in the  
 light, TAP medium in the dark, HS (minimal) medium in  
 ambient levels of CO2 and HS medium bubbled with 5% CO2.  
 PolyA mRNA was purified from each sample, pooled and cDNA  
 synthesized. The cDNA was directionally cloned into lambda  
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.  
 pBluescript II SK- plasmids were excised from the lambda  
 Zap clones by superinfection with ExAssist (Stratagene)  
 phage. The library was normalized using method 4 described  
 in Bonaldo et al (1996) Genome Research 6: 791-806."  
 BASE COUNT 95 a 202 c 107 g 145 t  
 ORIGIN

Query Match 79.2%; Score 19.8; DB 12; Length 549;

Best Local Similarity 91.3%; Pred. No. 3e+03;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATGGTGGCGGTGGGGGAGA 24

||||| ||||| ||||| ||||| |||||

DB 490 AAATGGTGGCGGTGGGGGAGA 468

||||| ||||| ||||| ||||| |||||

RESULT 11

BG501674

LOCUS

BG501674

DEFINITION

602548722F1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4634973 5',

mRNA sequence.

ACCESSION

BG501674

VERSION

BG501674.1 GI:13463191

BG501674 681 bp mRNA linear EST 27-MAR-2001

602548722F1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:4634973 5',

mRNA sequence.

BG501674

BG501674.1 GI:13463191





Search completed: January 19, 2003, 01:33:49  
Job time : 1572 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:06:14 ; Search time 1670 Seconds  
(without alignments)  
435.671 Million cell updates/sec

Title: US-09-674-593-9  
Perfect score: 25  
Sequence: 1 taggcgtgttggaaggtagcacca 25  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pin.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_man.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	25	100.0	1382	9	AF181722	AF181722 Homo sapi
C 2	25	100.0	4377	9	AF181720	AF181720 Homo sapi
C 3	25	100.0	120029	2	HSJ282H10	AL134672 Homo sapi
C 4	25	100.0	152966	9	AL359713	AL359713 Human DNA
C 5	20.4	81.6	147525	2	AC099072	AC099072 Rattus no
6	19.8	79.2	1270	14	TEN319582	AJ319582 Tick-born
7	19.8	79.2	1270	14	TEN319583	AJ319583 Tick-born
8	19.8	79.2	1270	14	TEN319584	AJ319584 Tick-born
9	19.8	79.2	1270	14	TEN319585	AJ319585 Tick-born
10	19.8	79.2	1270	14	TEN319586	AJ319586 Tick-born
11	19.8	79.2	1270	14	TI414703	AJ414703 Tick-born
12	19.8	79.2	1488	14	AF091005	AF091005 Tick-born
13	19.8	79.2	1488	14	AF091007	AF091007 Tick-born
14	19.8	79.2	1488	14	AF091009	AF091009 Tick-born
15	19.8	79.2	1488	14	AF091010	AF091010 Tick-born
16	19.8	79.2	1488	14	AF091011	AF091011 Tick-born
17	19.8	79.2	1488	14	AF091012	AF091012 Tick-born
18	19.8	79.2	1488	14	AF091014	AF091014 Tick-born
19	19.8	79.2	1488	14	AF091018	AF091018 Tick-born
20	19.8	79.2	1488	14	AF091020	AF091020 Tick-born
21	19.8	79.2	1488	14	FVIREP	X60286 Flavivirus
22	19.8	79.2	1488	14	LIVIEUV	X86785 Louping ill
23	19.8	79.2	1488	14	NEGENVGP	M94956 Negishi vir
24	19.8	79.2	1488	14	TBCVENV2	X76608 Tick-borne
25	19.8	79.2	1488	14	TBEVENV1	X76607 Tick-borne
26	19.8	79.2	2418	6	A02208	A02208 Flavivirus
C 27	19.8	79.2	2418	6	A02209	A02209 Flavivirus
28	19.8	79.2	2418	6	A02210	A02210 Flavivirus
29	19.8	79.2	11141	6	A48220	A48220 Sequence 19
30	19.8	79.2	11141	6	A57313	A57313 Sequence 19
31	19.8	79.2	11141	14	TEU27491	U27491 Tick-borne
32	19.8	79.2	11141	14	TEU27495	U27495 Tick-borne
C 33	19.8	79.2	171715	2	AC117028	AC117028 Rattus no
34	19.2	76.8	3166	10	MMCKITM	X65997 M.musculus
35	19.2	76.8	5098	6	AX195906	AX195906 Sequence
36	19.2	76.8	5098	10	MMCKIT	Y00864 Mouse c-kit
37	19.2	76.8	154995	2	AC115853	AC115853 Mus muscu
38	19.2	76.8	167960	2	AC027539	AC027539 Homo sapi
C 39	19.2	76.8	171734	9	AP003460	AP003460 Homo sapi
40	19.2	76.8	172344	2	AC011176	AC011176 Mus muscu
41	19.2	76.8	211316	9	AC009294	AC009294 Homo sapi
42	19.2	76.8	234285	2	AC105849	AC105849 Rattus no
43	18.8	75.2	110294	2	AC105640	AC105640 Rattus no
44	18.8	75.2	142690	2	AC102207	AC102207 Mus muscu
C 45	18.8	75.2	149218	2	AC023997	AC023997 Homo sapi

ALIGNMENTS

RESULT 1  
AF181722/c  
LOCUS AF181722 1382 bp mRNA linear PRI 10-JAN-2000  
DEFINITION Homo sapiens RU2AS (RU2) mRNA, complete cds.  
ACCESSION AF181722  
VERSION AF181722.1 GI:6684531  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1382)  
AUTHORS Van Den Eynde, B.J., Gaugler, B., Probst-Kepper, M., Michaux, L.,  
Devuyt, O., Lorge, F., Weynants, P., and Boon, T.  
TITLE A new antigen recognized by cytolytic T lymphocytes on a human

kidney tumor results from reverse strand transcription  
J. Exp. Med. 190 (12), 1793-1800 (1999)  
MEDLINE 20069887  
PUBMED 10601354  
REFERENCE 2 (bases 1 to 1382)  
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilotte,L.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
Avenue Hippocrate, 74, Brussels 1200, Belgium  
FEATURES  
source  
1. .1382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p22.1"  
1. .1382  
/gene="RU2"  
738. .992  
/gene="RU2"  
/codon\_start=1  
/product="RU2AS"  
/protein\_id="AAF23613.1"  
/db\_xref="GI:6684532"  
/translation="MDDAAPRVGVPVAVHKHALHDLGRLQVAGFGAAAHLPWP  
QLAASRRAPPLSQRPHTQAGSPPETNEKLTNPQVKEK"  
BASE COUNT 355 a 373 c 344 g 310 t  
ORIGIN  
  
Query Match 100.0%; Score 25; DB 9; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 0.09;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TAGCGCTGTTTGGAAAGGTTAGCACAC 25  
|||||  
Db 1304 TAGCGCTGTTTGGAAAGGTTAGCACAC 1280  
|||||  
  
RESULT 2  
AF181720  
LOCUS Homo sapiens RU2AS (RU2) gene, complete cds; and RU2S (RU2) gene,  
partial cds.  
DEFINITION AF181720  
ACCESSION AF181720.1 GI:6684526  
VERSION  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 4377)  
Van Den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L.,  
Devuyt,O., Lorge,F., Weynants,P. and Boon,T.  
A new antigen recognized by cytolytic T lymphocytes on a human  
kidney tumor results from reverse strand transcription  
J. Exp. Med. 190 (12), 1793-1800 (1999)  
MEDLINE 20069887  
PUBMED 10601354  
REFERENCE 2 (bases 1 to 4377)  
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilotte,L.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
Avenue Hippocrate, 74, Brussels 1200, Belgium  
FEATURES  
source  
1. .4377  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p22.1"  
1754. .>2348  
/gene="RU2"  
1754. .>2348  
/gene="RU2"  
/product="RU2S"

mRNA complement(<1912. .>2166)  
/product="RU2AS"  
CDS complement(1912. .2166)  
/codon\_start=1  
/product="RU2AS"  
/protein\_id="AAF23611.1"  
/db\_xref="GI:6684528"  
/translation="MDDAAPRVGVPVAVHKHALHDLGRLQVAGFGAAAHLPWP  
QLAASRRAPPLSQRPHTQAGSPPETNEKLTNPQVKEK"  
2056. .>2348  
/gene="RU2"  
/codon\_start=1  
/product="RU2S"  
/protein\_id="AAF23610.1"  
/db\_xref="GI:6684527"  
/translation="MSGSSARSSHLSPQVKSVLVYRNGDPFYAGRRVYIHEKKYSS  
EVFLKEVTGGVQAFPGAVRNITPTGRIRKLDIQSGGNFVAGGQAFKKL"  
BASE COUNT 1296 a 858 c 828 g 1380 t 15 others  
ORIGIN  
  
Query Match 100.0%; Score 25; DB 9; Length 4377;  
Best Local Similarity 100.0%; Pred. No. 0.083;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TAGCGCTGTTTGGAAAGGTTAGCACAC 25  
|||||  
Db 1600 TAGCGCTGTTTGGAAAGGTTAGCACAC 1624  
|||||  
  
RESULT 3  
HSJ282H10/c  
LOCUS Homo sapiens chromosome 6 clone RP1-282H10 map p22.1-22.3, \*\*\*  
DEFINITION SEQUENCING IN PROGRESS \*\*, 5 unordered pieces.  
ACCESSION AL132672  
VERSION AL132672.14 GI:14348905  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 120029)  
Garner,P.  
Direct Submission  
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Jun 12, 2001 this sequence version replaced gi:12331282.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: dj282H10  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: M13; M77815; 0% of reads  
Sequencing vector: plasmid; L08752; 99% of reads  
Chemistry: Dye-terminator ABI; 1% of reads  
Chemistry: Dye-terminator Big Dye; 98% of reads  
Consensus quality: 118224 bases at least Q40  
Consensus quality: 118449 bases at least Q30  
Consensus quality: 118634 bases at least Q20  
Insert size: 119629; sum-of-contigs  
Insert size: 117431; 9.3% error; agarose-fp  
Quality coverage: 18.41x in Q20 bases; sum-of-contigs Quality  
coverage: 19.36x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 23429: contig of 23429 bp in length  
 \* 23430 23529: gap of 100 bp  
 \* 23530 63600: contig of 40071 bp in length  
 \* 63601 63700: gap of 100 bp  
 \* 63701 100099: contig of 36399 bp in length  
 \* 100100 100199: gap of 100 bp  
 \* 100200 102366: contig of 2167 bp in length  
 \* 102367 102466: gap of 100 bp  
 \* 102467 120029: contig of 17563 bp in length.

## FEATURES

## source

1. 120029  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="p22.1-22.3"  
 /clone="RP1-282H10"  
 /clone\_lib="RPCI-1"  
 1. 23429  
 /note="assembly\_fragment:05837  
 clone\_end:SP6  
 vector\_side:left"  
 23530..63600  
 /note="assembly\_fragment:00423  
 fragment\_chain:1"  
 63701..100099  
 /note="assembly\_fragment:02685  
 fragment\_chain:1"  
 100200..102366  
 /note="assembly\_fragment:01789"  
 102467..120029  
 /note="assembly\_fragment:02341  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 36483 a 24840 c 24637 g 33666 t 403 others  
 ORIGIN

Query Match 100.0%; Score 25; DB 2; Length 120029;  
 Best Local Similarity 100.0%; Pred. No. 0.065;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTGGAAAGGCTAGGCACA 25

DB 22351 TAGGCTGTTGGAAAGGCTAGGCACA 22327

## RESULT 4

AL359713/c LOCUS 152966 bp DNA linear PRI 01-MAY-2001  
 DEFINITION Human DNA sequence from clone RP11-95P3 on chromosome 6, complete  
 sequence.

ACCESSION AL359713

VERSION AL359713.25 GI:13938809

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 152966)

AUTHORS Kimberley A.

TITLE Direct Submission

JOURNAL Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerquests@sanger.ac.uk

COMMENT On May 3, 2001 this sequence version replaced gi:13446455.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em, EMBL; Sw,  
 SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6  
 RP11-95P3 is from the library RPCI-11.1 constructed by the group of  
 Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone  
 RP11-95P3 It may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone RP11-95P3 is at 1 in this sequence. The  
 true left end of clone RP1-73M23 is at 152867 in this sequence. The  
 true right end of clone RP11-40E20 is at 17700 in this sequence.

## FEATURES

## source

1. 152966  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP11-95P3"  
 /clone\_lib="RPCI-11.1"  
 19..544  
 /note="match: GSS: Em:AZ517849"  
 24..379  
 /note="match: GSS: Em:B63526"  
 32..494  
 /note="match: GSS: Em:AQ315706"  
 70..589  
 /note="match: GSS: Em:AQ285677"  
 179..362  
 /note="MIR repeat: matches 65..262 of consensus"  
 521..582  
 /note="L2 repeat: matches 2683..2741 of consensus"  
 720..1022  
 /note="AluX repeat: matches 1..303 of consensus"  
 1327..1448  
 /note="AluJo/FRAM repeat: matches 181..298 of consensus"  
 2264..2575  
 /note="AluSq repeat: matches 1..311 of consensus"  
 3014..3296  
 /note="match: STS: Em:G27290"  
 3014..3093  
 /note="40 copies 2 mer ga 75% conserved"  
 5024..5477  
 /note="L1M1 repeat: matches 1012..1598 of consensus"  
 5581..5859  
 /note="AluSg repeat: matches 18..294 of consensus"  
 5986..6359  
 /note="L1MEC repeat: matches 2228..2259 of consensus"  
 6367..6515  
 /note="FLAM\_C repeat: matches 1..143 of consensus"  
 6971..7198  
 /note="L2 repeat: matches 1512..1735 of consensus"  
 8403..8838  
 /note="LTR7 repeat: matches 1..445 of consensus"  
 8799..9441  
 /note="match: GSS: Em:B54691"  
 complement(9393..9853)  
 /note="match: GSS: Em:AQ285817"  
 10311..10437  
 repeat\_region

```

/note="L2 repeat: matches 2612. .2747 of consensus"
10710. .11022
/note="AluJo repeat: matches 1. .309 of consensus"
12006. .12156
/note="L2 repeat: matches 2594. .2748 of consensus"
12182. .13011
/note="L2 repeat: matches 1374. .2344 of consensus"
13404. .13631
/note="114 copies 2 mer at 5% conserved"
13479. .13630
/note="38 copies 4 mer tata 60% conserved"
13785. .14820
/note="LIME repeat: matches 4766. .5798 of consensus"
14928. .16008
/note="L1PB3 repeat: matches 5106. .6149 of consensus"
16029. .16110
/note="L1PB3 repeat: matches 5054. .5133 of consensus"
16428. .16829
/note="L1MC3 repeat: matches 6189. .6603 of consensus"
16868. .17150
/note="AluX repeat: matches 1. .285 of consensus"
17246. .17991
/note="L1MC4 repeat: matches 6608. .7343 of consensus"
17992. .18675
/note="MER67C repeat: matches 1. .710 of consensus"
18676. .18729
/note="L1MC4 repeat: matches 7343. .7396 of consensus"
18731. .19458
/note="MER4D repeat: matches 232. .973 of consensus"
complement(18929. .19542)
/note="match: GSS: Em:AQ342961"
complement(19031. .19542)
/note="match: GSS: Em:AQ053552"
complement(19065. .19495)
/note="match: GSS: Em:AQ545892"
complement(19130. .19546)
/note="match: GSS: Em:B94529"
19265. .19495
/note="match: GSS: Em:AQ373351 Em:AQ375803
match: SFS: Em:G59266"
complement(19269. .19546)
/note="match: GSS: Em:AQ663911"
complement(19325. .19546)
/note="match: GSS: Em:AQ021494"
complement(19337. .19501)
/note="match: GSS: Em:AQ268095"
complement(19370. .19546)
/note="match: GSS: Em:AQ128461"
19566. .19739
/note="AluSg/x repeat: matches 126. .295 of consensus"
19740. .20100
/note="L1MD3 repeat: matches 7391. .7739 of consensus"
21121. .21429
/note="AluJo repeat: matches 1. .310 of consensus"
22589. .22728
/note="70 copies 2 mer aa 60% conserved"
22591. .22670
/note="20 copies 4 mer aaag 80% conserved"
22686. .22761
/note="19 copies 4 mer aggg 93% conserved"
23098. .23393
/note="AluJb repeat: matches 4. .298 of consensus"
23468. .23578
/note="L2 repeat: matches 2580. .2700 of consensus"
23980. .24292
/note="AluSg repeat: matches 2. .310 of consensus"
complement(27208. .27677)
/note="match: GSS: Em:AQ702871"
27819. .27957
/note="AluJb repeat: matches 163. .299 of consensus"
28701. .28930
/note="L2 repeat: matches 32. .301 of consensus"
29080. .29171

```

```

/note="MLT1H repeat: matches 433. .526 of consensus"
29303. .29711
/note="L2 repeat: matches 2323. .2710 of consensus"
29712. .29834
/note="MLT1G repeat: matches 29. .147 of consensus"
30140. .30547
/note="L2 repeat: matches 1806. .2252 of consensus"
complement(30720. .31207)
/note="match: SFS: Em:HS1J17"
30919. .31120
/note="MER58A repeat: matches 1. .208 of consensus"
32109. .32472
/note="THE1C repeat: matches 1. .371 of consensus"
33360. .33511
/note="MIR repeat: matches 98. .250 of consensus"
34031. .34331
/note="AluX repeat: matches 1. .297 of consensus"
34621. .34732
/note="MIR repeat: matches 120. .232 of consensus"
34888. .35007
/note="L2 repeat: matches 1836. .1959 of consensus"
35164. .35461
/note="AluSg repeat: matches 2. .298 of consensus"
35532. .35990
/note="L2 repeat: matches 2254. .2710 of consensus"
36631. .36790
/note="L2 repeat: matches 2342. .2501 of consensus"
37251. .37544
/note="AluX repeat: matches 1. .294 of consensus"
38446. .38493
/note="L2 copies 4 mer caca 75% conserved"
38467. .38492

Query Match 100.0%; Score 25; DB 9; Length 152966;
Best Local Similarity 100.0%; Pred. NO. 0.064;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTTGGAAAGGTTAGCACA 25
|||||
Db 54725 TAGGCTGTTTGGAAAGGTTAGCACA 54701

RESULT 5
AC099072/c
LOCUS Rattus norvegicus clone CH230-45P16, *** SEQUENCING IN PROGRESS
DEFINITION *** 59 unordered pieces.
ACCESSION AC099072
VERSION AC099072.3 GI:21723967
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 147525)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Blumage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,

```

```

Homs1, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louissege, H.,
Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapue, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mayhew, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wlarczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 147525)
Worley, K. C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 147525)
Worley, K. C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17957215.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFYF
Center clone name: CH230-45P16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 85085 bases at least Q40
Consensus quality: 93184 bases at least Q30
Consensus quality: 100226 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1262: contig of 1262 bp in length
* 1263 1362: gap of unknown length
* 1363 2404: contig of 1042 bp in length
* 2405 2504: gap of unknown length
* 2505 4050: contig of 1546 bp in length
* 4051 4150: gap of unknown length
* 4151 5791: contig of 1641 bp in length
* 5792 5891: gap of unknown length
* 5892 7405: contig of 1514 bp in length
*
7406 7505: gap of unknown length
7506 8640: contig of 1135 bp in length
8641 8740: gap of unknown length
8741 10160: contig of 1420 bp in length
10161 10260: gap of unknown length
10261 11621: contig of 1361 bp in length
11622 11721: gap of unknown length
11722 13393: contig of 1672 bp in length
13394 13493: gap of unknown length
13494 14576: contig of 1483 bp in length
14577 15076: gap of unknown length
15077 16398: contig of 1322 bp in length
16399 16498: gap of unknown length
16499 17708: contig of 1210 bp in length
17709 17808: gap of unknown length
17809 19041: contig of 1233 bp in length
19042 19141: gap of unknown length
19142 20366: contig of 1225 bp in length
20367 20466: gap of unknown length
20467 21686: contig of 1220 bp in length
21687 21786: gap of unknown length
21787 23061: contig of 1275 bp in length
23062 23161: gap of unknown length
23162 24536: contig of 1375 bp in length
24537 24636: gap of unknown length
24637 26086: contig of 1450 bp in length
26087 26186: gap of unknown length
26187 28301: contig of 2115 bp in length
28302 28401: gap of unknown length
28401 29921: contig of 1520 bp in length
29922 30021: gap of unknown length
30022 31486: contig of 1465 bp in length
31487 31586: gap of unknown length
31587 33258: contig of 1672 bp in length
33259 33358: gap of unknown length
33359 34769: contig of 1411 bp in length
34770 34869: gap of unknown length
34870 36962: contig of 2093 bp in length
36963 37062: gap of unknown length
37063 39041: contig of 1979 bp in length
39042 39141: gap of unknown length
39142 40252: contig of 1111 bp in length
40253 40352: gap of unknown length
40353 43547: contig of 3195 bp in length
43548 43647: gap of unknown length
43648 45760: contig of 2113 bp in length
45761 45860: gap of unknown length
45861 47866: contig of 2006 bp in length
47867 47966: gap of unknown length
47967 50477: contig of 2511 bp in length
50478 50577: gap of unknown length
50578 51957: contig of 1380 bp in length
51958 52057: gap of unknown length
52058 54347: contig of 2290 bp in length
54348 54447: gap of unknown length
54448 56934: contig of 2487 bp in length
56935 57034: gap of unknown length
57035 58741: contig of 1707 bp in length
58742 58841: gap of unknown length
58842 61254: contig of 2413 bp in length
61255 61354: gap of unknown length
61355 63297: contig of 1943 bp in length
63298 63397: gap of unknown length
63398 65668: contig of 2271 bp in length
65669 65768: gap of unknown length
65769 68380: contig of 2612 bp in length
68381 68480: gap of unknown length
68481 70587: contig of 2107 bp in length
70588 70687: gap of unknown length
70688 72922: contig of 2235 bp in length
72923 73022: gap of unknown length
73023 76477: contig of 3455 bp in length
76478 76577: gap of unknown length

```

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Tue Jan 21 10:17:06 2003

```
* 76578 77698: contig of 1121 bp in length
* 77699 77798: gap of unknown length
* 77799 81194: contig of 3396 bp in length
* 81194 81294: gap of unknown length
* 81294 83328: contig of 2034 bp in length
* 83328 83428: gap of unknown length
* 83428 86777: contig of 3349 bp in length
* 86777 86878: gap of unknown length
* 86878 88661: contig of 1784 bp in length
* 88661 88761: gap of unknown length
* 88761 91147: contig of 2386 bp in length
* 91147 91247: gap of unknown length
* 91247 95601: contig of 4354 bp in length
* 95601 95701: gap of unknown length
* 95701 99139: contig of 3437 bp in length
* 99139 99238: gap of unknown length
* 99238 103686: contig of 4448 bp in length
* 103686 103786: gap of unknown length
* 103786 107375: contig of 3589 bp in length
* 107375 107475: gap of unknown length
* 107475 111116: contig of 3641 bp in length
* 111116 111217: gap of unknown length
* 111217 114696: contig of 3480 bp in length
* 114696 114797: gap of unknown length
* 114797 119320: contig of 4524 bp in length

Query Match      81.6%; Score 20.4; DB 2; Length 147525;
Best Local Similarity 95.5%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGCTGTTTGGAAAGGTTAGC 22
|||||
Db 63554 TAGGCTGTTTGGAAAGGTTAGC 63533

RESULT 6
TEN319582
LOCUS
DEFINITION
TICK-borne encephalitis virus partial E gene for envelope protein,
genomic RNA, strain Latvia-11686.
ACCESSION
AJ319582.1 GI:18076914
VERSION
E gene; envelope protein.
KEYWORDS
TICK-borne encephalitis virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; tick-borne encephalitis virus group.
REFERENCE
1
Lundkvist,K., Vene,S., Golovljova,I., Mavtchoutko,V., Forsgren,M.,
Kalinina,V. and Plyusnin,A.
Characterization of tick-borne encephalitis virus from Latvia:
evidence for co-circulation of three distinct subtypes
J. Med. Virol. 65 (4), 730-735 (2001)
21611693
PUBMED
11745938
REFERENCE
2 (bases 1 to 1270)
Plyusnin,A.
Direct Submission
Submitted (22-MAY-2001) Plyusnin A., Department of Virology,
University of Helsinki, Haartman Institute, Haartmaninkatu 3, P.O.
Box 21, FIN-00014, FINLAND
Location/Qualifiers
1. .1270
/organism="Tick-borne encephalitis virus"
/virus
/isolate="Latvia-11686"
/db_xref="taxon:11084"
/lab_host="suckling white mice"
/country="Latvia"
1. .1270
/gene="E"
<1. .>1270
/gene="E"

gene
CDS

TEN319583
LOCUS
DEFINITION
TICK-borne encephalitis virus partial E gene for envelope protein,
genomic RNA, strain Latvia-8110.
ACCESSION
AJ319583.1 GI:18076916
VERSION
E gene; envelope protein.
KEYWORDS
TICK-borne encephalitis virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; tick-borne encephalitis virus group.
REFERENCE
1
Lundkvist,K., Vene,S., Golovljova,I., Mavtchoutko,V., Forsgren,M.,
Kalinina,V. and Plyusnin,A.
Characterization of tick-borne encephalitis virus from Latvia:
evidence for co-circulation of three distinct subtypes
J. Med. Virol. 65 (4), 730-735 (2001)
21611693
PUBMED
11745938
REFERENCE
2 (bases 1 to 1270)
Plyusnin,A.
Direct Submission
Submitted (22-MAY-2001) Plyusnin A., Department of Virology,
University of Helsinki, Haartman Institute, Haartmaninkatu 3, P.O.
Box 21, FIN-00014, FINLAND
Location/Qualifiers
1. .1270
/organism="Tick-borne encephalitis virus"
/virus
/isolate="Latvia-8110"
/db_xref="taxon:11084"
/lab_host="suckling white mice"
/country="Latvia"
1. .1270
/gene="E"
<1. .>1270
/gene="E"

gene
CDS

mat_peptide
BASE COUNT 361 a 270 c 377 g 262 t
ORIGIN
Query Match      79.2%; Score 19.8; DB 14; Length 1270;
Best Local Similarity 91.3%; Pred. No. 41;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGCTGTTTGGAAAGGTTAGCA 23
|||||
Db 274 TGGACTGTTTGGAAAGGTTAGCA 296

RESULT 7
TEN319583
LOCUS
DEFINITION
TICK-borne encephalitis virus partial E gene for envelope protein,
genomic RNA, strain Latvia-8110.
ACCESSION
AJ319583.1 GI:18076916
VERSION
E gene; envelope protein.
KEYWORDS
TICK-borne encephalitis virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Flavivirus; tick-borne encephalitis virus group.
REFERENCE
1
Lundkvist,K., Vene,S., Golovljova,I., Mavtchoutko,V., Forsgren,M.,
Kalinina,V. and Plyusnin,A.
Characterization of tick-borne encephalitis virus from Latvia:
evidence for co-circulation of three distinct subtypes
J. Med. Virol. 65 (4), 730-735 (2001)
21611693
PUBMED
11745938
REFERENCE
2 (bases 1 to 1270)
Plyusnin,A.
Direct Submission
Submitted (22-MAY-2001) Plyusnin A., Department of Virology,
University of Helsinki, Haartman Institute, Haartmaninkatu 3, P.O.
Box 21, FIN-00014, FINLAND
Location/Qualifiers
1. .1270
/organism="Tick-borne encephalitis virus"
/virus
/isolate="Latvia-8110"
/db_xref="taxon:11084"
/lab_host="suckling white mice"
/country="Latvia"
1. .1270
/gene="E"
<1. .>1270
/gene="E"

gene
CDS

/codon_start=2
/evidence-experimental
/product="envelope protein"
/protein_id="CAC85360.1"
/db_xref="GI:18076915"
/translation="TQGTTRVTLVLELGGCVTTAEGKPSMDVWLDIAYOENPAKTRE
YCLHAKLSDTKVAARCTMGTPATLAEEHGGTVCKRQSDRGWNGHGLFGKGSIVAC
VKACEAKKATGHVYDANKIVYTVKEPHTGDYVAANETHSGRTASTFTVSSSEKTIIL
TMGEYGDVSYLLCRVASGVDLAQTVILEDKTVEHLPTAWQVHRDWFNDLALPKHGEA
QNNNAERLVEFGAPHAVKMDVYNLGDTQTVLLKALAGVPAHTEGTYHLKSGHVTG
EVLGELKMKGLTYTCDKTKFTWKRAPTDSGHDITVYMEVTFSGTKFCRIPYRAVHG
SPDVNVAMLITPNPTIENGGGFTEMQLPPGDNIYYVVELSHQWFGKSSIGRVFQKI
KKGIERTVIGEHAWDFGAGGFLSSIGAV"
<1. .>1270
/gene="E"
/product="envelope protein"
/evidence-experimental 262 t
BASE COUNT 361 a 270 c 377 g 262 t
ORIGIN
```







[illegible]



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:44:49 ; Search time 57.5 Seconds  
(without alignments)  
193.856 Million cell updates/sec

Title: US-09-674-593-9  
Perfect score: 25  
Sequence: 1 tagcgtgttggaaggtagcac 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 393868 seqs, 222934149 residues

Total number of hits satisfying chosen parameters: 787736

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	18.2	72.8	731	10	US-09-833-381-1755
2	18	72.0	372	10	US-09-954-456-392
3	18	72.0	372	10	US-09-954-456-1357
4	18	72.0	372	10	US-09-954-456-2252
5	18	72.0	372	10	US-09-880-107-1971
c 6	17.6	70.4	3690	9	US-10-174-590-517
c 7	17.6	70.4	3690	9	US-10-176-758-517
c 8	17.6	70.4	3690	12	US-10-052-586-517
c 9	17	68.0	578	10	US-09-864-761-12703
c 10	16.8	67.2	300	9	US-09-796-692-7437
11	16.6	66.4	264	10	US-09-878-574-14812
12	16.6	66.4	362	10	US-09-878-574-3221
c 13	16.6	66.4	520	10	US-09-864-761-13048
14	16.6	66.4	1080	9	US-09-895-913A-115
c 15	16.6	66.4	15000	9	US-09-954-531-175
16	16.2	64.8	231	9	US-10-046-935-1620
17	16.2	64.8	2124	10	US-09-878-178-1620
c 18	16.2	64.8	2124	10	US-09-815-925-8
19	16.2	64.8	33795	10	US-09-880-107-2184

ALIGNMENTS

RESULT 1

US-09-833-381-1755/c  
; Sequence 1755, Application US/09833381  
; Patent No. US20020132090A1  
; GENERAL INFORMATION:  
; APPLICANT: Robison, Keith E.  
; TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs  
; FILE REFERENCE: 5800-119  
; CURRENT APPLICATION NUMBER: US/09/833,381  
; CURRENT FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 09/516,448  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 2050  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1755  
; LENGTH: 731  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(731)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-833-381-1755

Query Match 72.8%; Score 18.2; DB 10; Length 731;  
Best Local Similarity 83.3%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AGGCTGTTTGGAAAGGTAGCAC 25  
| | | | | | | | | | | | | | | | | |  
Db 173 AGGCTGGTGGGAAGGTGGCNCA 150

RESULT 2

US-09-954-456-392  
; Sequence 392, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76

Sequence 31909, A  
Sequence 3860, Ap  
Sequence 158, App  
Sequence 8299, App  
Sequence 11427, A  
Sequence 710, Appl  
Sequence 55, Appl  
Sequence 5452, Ap  
Sequence 404, App  
Sequence 616, App  
Sequence 10, Appl  
Sequence 204, App  
Sequence 93, Appl  
Sequence 18, Appl  
Sequence 134, App  
Sequence 1, Appl  
Sequence 718, App  
Sequence 43, Appl  
Sequence 53, Appl  
Sequence 54, Appl  
Sequence 4652, Ap  
Sequence 203, App  
Sequence 3386, Ap  
Sequence 3327, Ap  
Sequence 670, App  
Sequence 695, App

; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 392  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-392

Query Match 72.0%; Score 18; DB 10; Length 372;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTGGAAAGGG 18  
|||||  
Db 279 TAGGCTGTTGGAAAGGG 296

RESULT 3  
US-09-954-456-1357  
; Sequence 1357, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27

; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1357  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-1357

Query Match 72.0%; Score 18; DB 10; Length 372;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTGGAAAGGG 18  
|||||  
Db 279 TAGGCTGTTGGAAAGGG 296

RESULT 4  
US-09-954-456-2252  
; Sequence 2252, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2252  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-456-2252

Query Match 72.0%; Score 18; DB 10; Length 372;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTGGAAAGGG 18  
|||||  
Db 279 TAGGCTGTTGGAAAGGG 296

RESULT 5  
US-09-880-107-1971

; Sequence 1971, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1971  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 H62838  
; NAME/KEY: unsure  
; LOCATION: (1)..(372)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-880-107-1971

Query Match 72.0%; Score 18; DB 10; Length 372;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TAGGCTGTTGGAAAGG 18  
Db 279 TAGGCTGTTGGAAAGG 296  
|||||

## RESULT 6

US-10-174-590-517/c  
; Sequence 517, Application US/10174590  
; Publication No. US20030008352A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C42  
; CURRENT APPLICATION NUMBER: US/10/174,590  
; CURRENT FILING DATE: 2002-06-18  
; Prior application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 517  
; LENGTH: 3690  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-174-590-517

Query Match 70.4%; Score 17.6; DB 9; Length 3690;  
Best Local Similarity 83.3%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AGGCTGTTGGAAAGGTTAGCAC 25  
Db 2135 AGGCTGTTGGAAAGGTTAGCAC 2112  
|||||

## RESULT 7

US-10-176-758-517/c  
; Sequence 517, Application US/10176758  
; Publication No. US20030008353A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C104  
; CURRENT APPLICATION NUMBER: US/10/176,758  
; CURRENT FILING DATE: 2002-06-21  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 517  
; LENGTH: 3690  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-176-758-517

Query Match 70.4%; Score 17.6; DB 9; Length 3690;  
Best Local Similarity 83.3%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AGGCTGTTGGAAAGGTTAGCAC 25  
Db 2135 AGGCTGTTGGAAAGGTTAGCAC 2112  
|||||

## RESULT 8

US-10-052-586-517/c  
; Sequence 517, Application US/10052586  
; Patent No. US20020127584A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C1  
; CURRENT APPLICATION NUMBER: US/10/052,586  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21

[illegible]

; PRIOR APPLICATION NUMBER: 60/088826  
; PRIOR FILING DATE: 1998-06-10  
; PRIOR APPLICATION NUMBER: 60/088861  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088863  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/088876  
; PRIOR FILING DATE: 1998-06-11  
; PRIOR APPLICATION NUMBER: 60/089090  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089105  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089512  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089514  
; PRIOR FILING DATE: 1998-06-16  
; PRIOR APPLICATION NUMBER: 60/089538  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089598  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089653  
; PRIOR FILING DATE: 1998-06-17  
; PRIOR APPLICATION NUMBER: 60/089908

Query Match 70.4%; Score 17.6; DB 12; Length 3690;  
Best Local Similarity 83.3%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AGCGTGTGGAAAGGTTAGCACCA 25  
||||| | ||||| || ||  
Db 2135 AGCGTGTGGAAAGGTTAGCCCA 2112

## RESULT 9

US-09-864-761-12703/c  
; Sequence 12703, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 12703  
; LENGTH: 578  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC023114.3  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8  
US-09-864-761-12703

Query Match 68.0%; Score 17; DB 10; Length 578;  
Best Local Similarity 80.0%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TAGCGTGTGGAAAGGTTAGCACCA 25  
| | | | | | | | | | | | | | | |  
Db 296 TGGACTGTTTGAAGGGGATACA 272

## RESULT 10

US-09-796-692-7437/c  
; Sequence 7437, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane  
; APPLICANT: Algate, Paul A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597







**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:34:34 ; Search time 1569 Seconds  
(without alignments)  
258.054 Million cell updates/sec

Title: US-09-674-593-9  
Perfect score: 25  
Sequence: 1 taggcgtgttgaaaggtagcaca 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estnu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hct:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hct3:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: gb\_gss:\*
  - 18: em\_gss\_hum:\*
  - 19: em\_gss\_inv:\*
  - 20: em\_gss\_pln:\*
  - 21: em\_gss\_vrt:\*
  - 22: em\_gss\_fun:\*
  - 23: em\_gss\_mam:\*
  - 24: em\_gss\_mus:\*
  - 25: em\_gss\_other:\*
  - 26: em\_gss\_pro:\*
  - 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23	92.0	879	14	B0953400 AGENCOURT
C 2	22	88.0	1045	12	BG749180 602708139
C 3	19.2	76.8	491	17	TA191F11Q
4	19.2	76.8	592	17	A0649802
5	19.2	76.8	603	17	A0949417
6	19.2	76.8	624	10	AW323372

7	19.2	76.8	698	13	BG976420
8	19.2	76.8	700	13	BI081374
9	19.2	76.8	751	13	BI105298
10	19.2	76.8	789	13	BI081897
C 11	19.2	76.8	790	13	BM230522
12	19.2	76.8	793	13	BI691212
13	19.2	76.8	839	12	BF143945
14	19.2	76.8	840	9	AU050911
15	19.2	76.8	885	14	BQ919757
16	19.2	76.8	891	12	BF139543
17	19.2	76.8	902	10	BE282881
18	19.2	76.8	1047	13	BI559229
19	19.2	76.8	5174	11	BC026713
20	18.6	74.4	229	10	BB057249
21	18.6	74.4	278	10	BB195421
C 22	18.6	74.4	422	14	BQ512701
23	18.6	74.4	465	9	AI477839
C 24	18.6	74.4	515	17	AZ788409
25	18.6	74.4	542	14	BQ259600
C 26	18.6	74.4	553	13	BI743491
27	18.6	74.4	601	10	AW566609
28	18.6	74.4	619	14	BM776459
29	18.6	74.4	650	14	BM776416
C 30	18.6	74.4	766	9	AF155036
31	18.6	74.4	856	13	BG985672
32	18.6	74.4	963	12	BF662899
C 33	18.2	72.8	241	10	AV358246
C 34	18.2	72.8	395	17	AQ332115
C 35	18.2	72.8	559	14	BQ170080
C 36	18.2	72.8	618	10	BE211330
C 37	18.2	72.8	678	17	AZ977273
C 38	18.2	72.8	795	10	BE641995
C 39	18.2	72.8	952	12	BF143993
40	18.2	72.8	1009	14	BQ957654
41	18	72.0	372	14	H62838
42	18	72.0	675	17	AG165617
43	17.8	71.2	140	10	AW863542
44	17.8	71.2	397	10	AW148538
C 45	17.8	71.2	433	10	BE790491

ALIGNMENTS

RESULT 1  
BQ953400/c  
LOCUS AGENCOURT\_8784199 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:6376296  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ953400  
VERSION BQ953400.1 GI:22368878  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 879)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM2558 row: e column: 01  
High quality sequence stop: 690.  
Location/Qualifiers  
1. .879  
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6376296"
/clone_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library. |"
BASE COUNT    194 a  275 c  236 g  173 t
ORIGIN
Query Match    92.0%; Score 23; DB 14; Length 879;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  3  GGCTGTTGGAAAGGGTAGCACA 25
      |||||||||||||||||||
Db  778 GGCTGTTGGAAAGGGTAGCACA 756

RESULT 2
BG749180/c
LOCUS          1045 bp  mRNA  linear  EST 15-MAY-2001
DEFINITION    602708139F1 NTH_MGC_43 Homo sapiens cDNA clone IMAGE:4844744 5',
              mRNA sequence.
ACCESSION     BG749180
VERSION       BG749180.1 GI:14059833
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 1045)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgabbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Prepared by: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLCM1682 row: n column: 09
               High quality sequence stop: 822.
               Location/Qualifiers
                 1. .1045
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:4844744"
                   /clone_lib="NIH_MGC_43"
                   /tissue_type="normal pigmented retinal epithelium"
                   /lab_host="DH10B (phage-resistant)"
                   /notes="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                   EcoRI; cDNA made by oligo-dT priming. Directionally
                   cloned into EcoRI/XhoI sites using the following 5'
                   adaptor: GGACGAG(G). Library constructed by Ling Hong
                   in the laboratory of Gerald M. Rubin (University of
                   California, Berkeley) using ZAP-cDNA synthesis kit
                   (Stratagene) and Superscript II RT (Life Technologies).
                   Note: this is a NIH_MGC Library. |"
BASE COUNT    245 a  322 c  300 g  178 t
ORIGIN
Query Match    88.0%; Score 22; DB 12; Length 1045;

```

```

Best Local Similarity 100.0%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  4  GCTGTTGGAAAGGGTAGCACA 25
      |||||||||||||||||||
Db  783 GCTGTTGGAAAGGGTAGCACA 762

RESULT 3
TA191F11Q
LOCUS          491 bp  DNA  linear  GSS 13-DEC-2000
DEFINITION    T. brucei sheared genomic DNA clone 191f11, reverse sequence,
              genomic survey sequence.
ACCESSION     AL477353
VERSION       AL477353.1 GI:11841379
KEYWORDS      GSS.
SOURCE        Trypanosoma brucei.
ORGANISM      Trypanosoma brucei
               Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
               Trypanosoma.
REFERENCE     1 (bases 1 to 491)
AUTHORS      Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
               Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
               Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE        Direct Submission
JOURNAL       Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
              project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
              Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
              nh@sanger.ac.uk
COMMENT       Constructed at the Institute for Genomic Research (TIGR),
              Rockville, MD. Genomic DNA isolated from a cloned population of
              Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
              to give a tight size distribution (
              4 kb). The v + i method used for the library construction is
              described in detail in Smith, H. and Venter, J.C. (Making small
              insert libraries for whole genome shotgun sequencing projects. In
              Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
              Barrell, Oxford University Press, 1999).
              Email: nelsayed@tigr.org
              Details of T. brucei sequencing at the Sanger Centre are available
              at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES       Location/Qualifiers
               source
                 1. .491
                   /organism="Trypanosoma brucei"
                   /strain="TREU927"
                   /db_xref="taxon:5691"
                   /clone="191f11"
BASE COUNT    157 a  129 c  125 g  80 t
ORIGIN
Query Match    76.8%; Score 19.2; DB 17; Length 491;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  2  AGCTCTTTGGAAGGGTAGCACA 25
      |||||||||||||||||||
Db  455 AGTCTGTTGGAATGGGTAGCACA 478

RESULT 4
AQ649802
LOCUS          592 bp  DNA  linear  GSS 22-JUN-1999
DEFINITION    Sheared DNA-16K8.TR Sheared DNA Trypanosoma brucei genomic clone
              Sheared DNA-16K8, DNA sequence.
ACCESSION     AQ649802
VERSION       AQ649802.1 GI:5142988
KEYWORDS      GSS.
SOURCE        Trypanosoma brucei.
ORGANISM      Trypanosoma brucei
               Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
               Trypanosoma.
REFERENCE     1 (bases 1 to 592)
AUTHORS      El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,

```

Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.  
 Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library  
 Unpublished (1999)  
 Other\_GSSs: Sheared DNA-16K8.TF  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mdb/tbdb/>.  
 Seq primer: M13-Reverse  
 Class: shotgun.

#### FEATURES

Location/Qualifiers  
 1. .592

/organism="Trypanosoma brucei"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone\_lib="Sheared DNA-16K8"  
 /clone\_lib="Sheared DNA"  
 /note="Vector: pUC18; Site\_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 236 a 124 c 148 g 84 t

ORIGIN

Query Match 76.8%; Score 19.2; DB 17; Length 592;

Best Local Similarity 87.5%; Pred. No. 3.4e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGCTGTTTGGAAAGGGTAGCACA 25

II IIIIIIIIIIIIIIIIIIIII

Db 229 AGTCTGTTGGAATGGTAGCACA 252

RESULT 5

AQ949417

LOCUS

DEFINITION Sheared DNA-31K11.TR Sheared DNA trypanosoma brucei genomic clone

Accession AQ949417

Version AQ949417.1 GI:6772682

Keywords GSS.

Source Trypanosoma brucei.

Organism Trypanosoma brucei

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 603)

Authors El-Sayed,N., Zhao,S., Gill,H., Sui,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.

Title Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library

Journal Unpublished (1999)

Other\_GSSs: Sheared DNA-31K11.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

FEATURES  
 source  
 1. .603

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db\_xref="taxon:5691"

/clone\_lib="Sheared DNA-31K11"

/clone\_lib="Sheared DNA"

/note="Vector: pUC18; Site\_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 194 a 151 c 158 g 100 t

ORIGIN

Query Match 76.8%; Score 19.2; DB 17; Length 603;

Best Local Similarity 87.5%; Pred. No. 3.4e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGCTGTTTGGAAAGGGTAGCACA 25

II IIIIIIIIIIIIIIIIIIIII

Db 575 AGTCTGTTGGAATGGTAGCACA 598

RESULT 6

AW323372

LOCUS

DEFINITION

AW323372 624 bp mRNA linear EST 26-JAN-2000  
 similar to gb:X65997 M.musculus c-kit mRNA for truncated tyrosine-kinase (MOUSE);, mRNA sequence.

Accession AW323372

Version AW323372.1 GI:6757397

Keywords EST.

Source house mouse.

Organism Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Reference 1 (bases 1 to 624)

Authors NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

Title National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Journal Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)

MG1:1027212

Seq primer: -40RP from Gibco

High quality sequence stop: 430.

Location/Qualifiers

1. .624

/organism="Mus musculus"

FEATURES

source

```

/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:2646760"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT      159 a 128 c 154 g 183 t
ORIGIN
Query Match      76.8%; Score 19.2; DB 10; Length 624;
Best Local Similarity 87.5%; Pred. No. 3.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGCTGTTTGAAGGGTAGCACA 25
|||||
Db 236 AGGCTGTTTGAAGGGTAGCACA 259

RESULT 7
BG976420
LOCUS
DEFINITION
602846216F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4976882 5',
mRNA sequence.
ACCESSION
BG976420
VERSION
BG976420.1 GI:14364057
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10970 row: p column: 03
High quality sequence stop: 695.
FEATURES
Location/Qualifiers
source
1..698
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4976882"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      183 a 144 c 181 g 189 t
ORIGIN
Query Match      76.8%; Score 19.2; DB 13; Length 698;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGCTGTTTGAAGGGTAGCACA 25
|||||
Db 236 AGGCTGTTTGAAGGGTAGCACA 259

RESULT 8
BI081374
LOCUS
DEFINITION
602879438F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5010968 5',
mRNA sequence.
ACCESSION
BI081374
VERSION
BI081374.1 GI:14499704
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1059 row: 1 column: 09
High quality sequence stop: 686.
FEATURES
Location/Qualifiers
source
1..700
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5010968"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT      178 a 138 c 185 g 199 t
ORIGIN
Query Match      76.8%; Score 19.2; DB 13; Length 700;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGCTGTTTGAAGGGTAGCACA 25
|||||
Db 511 AGGCTGTTTGAAGGGTAGCACA 534

RESULT 9
BI105298
LOCUS
DEFINITION
602893587F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5038677 5',
mRNA sequence.
ACCESSION
BI105298
VERSION
BI105298.1 GI:14556191
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

```

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM1106 row: n column: 22  
 High quality sequence stop: 741.

FEATURES  
source

Location/Qualifiers  
 1. .751  
 /organism="Mus musculus"  
 /strain="CZECH II"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5038677"  
 /clone\_lib="NCI\_CGAP\_Lu29"  
 /tissue\_type="spontaneous tumor, metastatic to mammary.  
 Stem cell origin."  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT  
ORIGIN

213 a 145 c 200 g 193 t

## Query Match

Best Local Similarity 76.8%; Score 19.2; DB 13; Length 751;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

## QY

2 AGGCTGTTTGGAAAGGTTAGCAC 25

||||| ||||| ||||| |||||

## Db

716 AGGCTGTTTGGAAAGGTTAGCAC 739

## RESULT 10

BI081897

LOCUS

602877120F1 NCI\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5008924 5',

mRNA sequence.

BI081897

BI081897.1 GI:14500227

EST.

house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 789)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM1054 row: g column: 05

High quality sequence stop: 725.

FEATURES  
source

Location/Qualifiers  
 1. .789  
 /organism="Mus musculus"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5008924"  
 /clone\_lib="NCI\_CGAP\_Mam2"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 months"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 184 a 212 c 186 g 207 t

## ORIGIN

Query Match 76.8%; Score 19.2; DB 13; Length 789;

Best Local Similarity 87.5%; Pred. No. 3.8e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

## QY

2 AGGCTGTTTGGAAAGGTTAGCAC 25

||||| ||||| ||||| |||||

## Db

142 AGGCTGTTTGGAAAGGTTAGCAC 165

## RESULT 11

BM230522/c

LOCUS

BM230522 790 bp mRNA linear EST 31-JAN-2002

K0296G07-3 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus

musculus cDNA clone K0296G07 3', mRNA sequence.

BM230522

ACCESSION

VERSION

KEYWORDS

EST.

BM230522.1 GI:17793765

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 790)

Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Stagg, C.A.,

Martin, P., Alba, K., Tanaka, T. and Ko, M.S.H.

Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library

(Long)

Unpublished (2001)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: [cdna@gsun.grc.nia.nih.gov](mailto:cdna@gsun.grc.nia.nih.gov)

Plate: K0296 row: G column: 07

Seq primer: -21M13 Forward

High quality sequence stop: 790

POLYA=Yes

Location/Qualifiers

1. .790

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="niaEST:K0296G07-3"

/db\_xref="taxon:10090"

/clone="K0296G07"

/clone\_lib="NIA Mouse Unfertilized Egg cDNA Library (Long

)"

/tissue\_type="Unfertilized Egg"

/lab\_host="DH10B"

/note="Vector: pSPORT1 (Invitrogen); Site\_1: SalI; Site\_2:

NotI; Mouse cDNA project by the Laboratory of Genetics,

National Institute on Aging (NIA), Intramural Research

Program, NIH (<http://igsun.grc.nia.nih.gov/cDNA>). This is

a long-transcript enriched cDNA library (Ref. Genome Res.

11: 1553-1558 (2001). [PMID: 11541191]. Total RNAs were

extracted from a pool of 1488 unfertilized eggs.

Double-stranded cDNAs were synthesized with an Oligo(dT)

primer [Invitrogen:

5'-pGACTAGTCTAGATCGGAGCGCGCTTTT-3'],

treated with T4 DNA polymerase, and purified by

ethanol-precipitation. The cDNAs were ligated to

lone-linker LL-SalI, purified by phenol/chloroform, and

separated from free linkers by Centricon 100. Then, the

cDNAs were amplified by long-range high fidelity PCR using

Ex Taq polymerase (Takara) with a primer Sal4-S. The

products were purified by phenol/chloroform and Centricon

100. The cDNAs were digested with SalI and NotI enzymes

and cloned into SalI/NotI site of pSPORT1 plasmid vector.  
The DH10B E. coli host was transformed with the ligation  
mixture by the standard chemical method. The average  
insert size is about 2.5 kb. The library was constructed  
by Yulan Piao (NIA)."

BASE COUNT 218 a 210 c 155 g 207 t  
ORIGIN

Query Match 76.8%; Score 19.2; DB 13; Length 790;  
Best Local Similarity 87.5%; Pred. No. 3.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AGGCTGTTTGGAAAGGTAGCACAC 25  
||||| ||||| ||||| ||||| |||||

Db 181 AGGCTGTTTGGAAAGGTACCAC 158

## RESULT 12

BI691212

LOCUS 793 bp mRNA linear EST 18-SEP-2001  
DEFINITION 603311015F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5347011 5',  
mRNA sequence.

ACCESSION BI691212

VERSION BI691212.1 GI:15653841

KEYWORDS EST.

SOURCE house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 793)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1881 row: n column: 04

High quality sequence stop: 777.

Location/Qualifiers

## FEATURES

source

1..793

/organism="Mus musculus"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone\_lib="IMAGE:5347011"

/clone\_lib="NCI\_CGAP\_Mam6"

/sex="female, virgin"

/tissue\_type="infiltrating ductal carcinoma"

/dev\_stage="5 months"

/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;

Site\_2: NotI; Cloned unidirectionally. Primer: Oligo 3'

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 198 a 170 c 203 g 222 t

## ORIGIN

Query Match 76.8%; Score 19.2; DB 13; Length 793;

Best Local Similarity 87.5%; Pred. No. 3.8e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AGGCTGTTTGGAAAGGTAGCACAC 25

||||| ||||| ||||| ||||| |||||

Db 532 AGGCTGTTTGGAAAGGTACCAC 555

## RESULT 13

BF143945

## LOCUS

DEFINITION

601786489F1 NCI\_CGAP\_Lu30 Mus musculus cDNA clone IMAGE:4014238 5',

mRNA sequence.

ACCESSION BF143945

VERSION BF143945.1 GI:10982985

KEYWORDS EST.

SOURCE house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 839)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9258 row: a column: 23

High quality sequence stop: 590.

Location/Qualifiers

## FEATURES

source

1..839

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone\_lib="IMAGE:4014238"

/clone\_lib="NCI\_CGAP\_Lu30"

/tissue\_type="tumor, metastatic to mammary"

/lab\_host="DH10B"

/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; transgenic model WNT-1, expression driven by

MMTV-I-TR enhancer; Cloned unidirectionally. Primer: Oligo

3' Library constructed by Life Technologies.

Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 215 a 179 c 207 g 238 t

## ORIGIN

Query Match 76.8%; Score 19.2; DB 12; Length 839;

Best Local Similarity 87.5%; Pred. No. 3.9e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AGGCTGTTTGGAAAGGTAGCACAC 25

||||| ||||| ||||| ||||| |||||

Db 338 AGGCTGTTTGGAAAGGTACCAC 361

## RESULT 14

AU050911

LOCUS

DEFINITION

AU050911 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-1202

5', mRNA sequence.

ACCESSION AU050911

VERSION AU050911.1 GI:4433920

KEYWORDS EST.

SOURCE house mouse.

## ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 840)

Sasaki, M., Suzuki, Y., Watanabe, M., Imai, J., Shibui, A., Yoshida, K.,

Hata, H., Yamaguchi, R., Tateyama, S. and Sugano, S.

Construction of mouse full length-enriched cDNA libraries by

oligo-capping

Unpublished (1999)

Contact: Katsuyuki Hashimoto

Division of Genetic Resources

National Institute of Infectious Diseases

23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan



Email: khashi@nih.gov  
URL: http://www.nih.gov/yoken/genbank/  
Seq primer: 5' end primer: CTTCTGCTCTAAAGCTCCG  
POLYA-No. Location/Qualifiers

FEATURES

1. .840  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="MNCB-1202"  
/clone\_lib="Sugano mouse brain mncb"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="TOP10"  
/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA was primed with an oligo(dT) primer ATGCGCCCTTTTCTTTTCTTTT; double-stranded cDNA was ligated to a DraIII adaptor (TGTGGCCTACTGG), digested and cloned into distinct DraIII sites of the pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer used for sequencing: 5' end primer [CTTCTGCTCTAAAGCTCCG], 3' end primer [CGACCTGCAGCTCGACACA]"

BASE COUNT 157 a 198 c 129 g 252 t 104 others  
ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 840;  
Best Local Similarity 87.5%; Pred. No. 3.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGCTGTTTGGAAAGGTTAGCACA 25  
|||||  
DB 119 AGGCTGTTTGGAAAGGTTAGCACA 142

RESULT 15

BQ919757  
LOCUS BQ919757 885 bp mRNA linear EST 20-AUG-2002  
DEFINITION AGENCOURT\_8819013 NCI\_CGAP\_Mam2 Mus musculus cDNA clone  
IMAGE:6438574 5', mRNA sequence.  
ACCESSION BQ919757  
VERSION BQ919757.1 GI:22334455  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 885)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Cloning Distribution: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLMI3957 row: c column: 23  
High quality sequence stop: 696.  
Location/Qualifiers  
1. .885  
/organism="Mus musculus"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6438574"  
/clone\_lib="NCI\_CGAP\_Mam2"  
/tissue\_type="tumor, biopsy sample"

FEATURES

1. .885  
/organism="Mus musculus"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6438574"  
/clone\_lib="NCI\_CGAP\_Mam2"  
/tissue\_type="tumor, biopsy sample"

/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: Sali;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"  
BASE COUNT 216 a 195 c 227 g 244 t 3 others  
ORIGIN  
Query Match 76.8%; Score 19.2; DB 14; Length 885;  
Best Local Similarity 87.5%; Pred. No. 4e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AGGCTGTTTGGAAAGGTTAGCACA 25  
|||||  
DB 520 AGGCTGTTTGGAAAGGTTAGCACA 543  
Search completed: January 19, 2003, 01:33:52  
Job time : 1572 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 23:06:14 ; Search time 1670 seconds  
(without alignments)  
435.671 Million cell updates/sec

Title: US-09-674-593-8

Perfect score: 25

Sequence: 1 taaatgggtggcggtggggagac 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pin.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	25	100.0	1287	9	BC014954	BC014954 Homo sapi
c 2	25	100.0	1382	9	AF181722	AF181722 Homo sapi
c 3	25	100.0	4377	9	AF181720	AF181720 Homo sapi
c 4	25	100.0	120029	2	HSJ282H10	AL132672 Homo sapi
c 5	25	100.0	152966	9	AL359713	AL359713 Human DNA
c 6	20.8	83.2	98724	2	AC115426	AL15426 Rattus no
c 7	20.8	83.2	128796	9	AL157396	AL157396 Human DNA
c 8	20.8	83.2	164773	2	AC091543	AC091543 Felis cat
c 9	20.4	81.6	168203	2	AC117062	AC117062 Rattus no
c 10	20.2	80.8	977	14	TAHYNAS	268497 Tahyna viru
c 11	20.2	80.8	70905	2	AC119906	AC119906 Mus muscu
c 12	20.2	80.8	180252	2	AC023909	AC023909 Mus muscu
c 13	20	80.0	126441	2	AC095226	AC095226 Rattus no
c 14	19.8	79.2	12141	6	AR129606	AR129606 Sequence
c 15	19.8	79.2	12141	10	AF009605	AF009605 Mus muscu
c 16	19.8	79.2	45562	9	AL358512	AL358512 Human DNA
c 17	19.8	79.2	87229	2	AC123538	AC123538 Sminthops
c 18	19.8	79.2	108299	10	AL606747	AL606747 Mouse DNA
c 19	19.8	79.2	147146	9	AC006058	AC006058 Homo sapi
c 20	19.8	79.2	154383	2	AC097312	AC097312 Rattus no
c 21	19.8	79.2	156075	2	AL837509	AL837509 Mus muscu
c 22	19.8	79.2	166445	2	AC128745	AC128745 Rattus no
c 23	19.8	79.2	167632	2	AC068604	AC068604 Mus muscu
c 24	19.8	79.2	176129	2	AC073222	AC073222 Homo sapi
c 25	19.8	79.2	185479	2	AC117092	AC117092 Rattus no
c 26	19.8	79.2	195966	2	AC102785	AC102785 Mus muscu
c 27	19.8	79.2	244931	2	AC125153	AC125153 Mus muscu
c 28	19.4	77.6	105581	2	AC108448_3	Continuation (4 of
c 29	19.4	77.6	110000	2	AC108448_0	AC108448 Homo sapi
c 30	19.4	77.6	110000	2	AC108448_1	Continuation (2 of
c 31	19.4	77.6	110000	2	AC108448_2	Continuation (3 of
c 32	19.2	76.8	204	14	SSH33VC	ML1782 Snowshoe ha
c 33	19.2	76.8	976	14	MBU31989	U31989 Morro Bay v
c 34	19.2	76.8	976	14	TVU47142	U47142 Tahyna viru
c 35	19.2	76.8	978	14	CEU12797	U12797 California
c 36	19.2	76.8	978	14	CEU12800	U12800 California
c 37	19.2	76.8	980	14	BLCSA	M19420 Germiston v
c 38	19.2	76.8	982	14	SSH5	J02390 Snowshoe ha
c 39	19.2	76.8	984	14	AF528167	AF528167 La Crosse
c 40	19.2	76.8	984	14	BLCSRNA	K00610 La Crosse v
c 41	19.2	76.8	1077	14	BLNSA	X73468 Lumbo virus
c 42	19.2	76.8	27555	2	AC012903	AC012903 Drosophil
c 43	19.2	76.8	36654	9	AF315943	AF315943 Homo sapi
c 44	19.2	76.8	44523	2	AC095810	AC095810 Rattus no
c 45	19.2	76.8	55590	9	AC073128	AC073128 Homo sapi

ALIGNMENTS

RESULT 1  
BC014954/c  
LOCUS BC014954 1287 bp mRNA linear PRI 04-OCT-2001  
DEFINITION Homo sapiens, clone MGC:22980 IMAGE:4874845, mRNA, complete cds.  
ACCESSION BC014954  
VERSION BC014954.1 GI:15928979  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1287)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (01-OCT-2001) National Institutes of Health, Mammalian

Pred. No. is the number of results predicted by chance to have a

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

BC Cancer Agency: Genome Sequence Centre,  
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 34 Row: m Column: 13

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7706690.

FEATURES  
source  
1. .1287  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:22980 IMAGE:4874845"  
/tissue\_type="Eye, normal, pigmented retinal epithelium"  
/clone\_lib="NIH\_MGC\_43"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
181. .984  
/codon\_start=1  
/product="Unknown (protein for MGC:22980)"  
/protein\_id="AAH14954.1"  
/db\_xref="GI:15928980"  
/translation="MCGAVSAGEDNDLDNLKEAQYINTERVEQAFRAIDRGDYILE  
GYNDNAYKDLAWKHGNIHLSAFICIVSEVMALKLQPLGSLNLGSGTGLYLSTMVGLIL  
GPGFNGHLELSDVVEYAKLEISFKNSDFDKRRSRALGAVPLGQRGLPRSS  
ELRRPAGKMSGSSARSHLSQPVVKSIVYRNGDPFVAGRRVVIHEKKVSVFEVFLK  
EVTGGVQAPFGAVRNITPTGTGRTKRLQIQSGNYYVAGGQEAFFKLK"

BASE COUNT 343 a 281 c 359 g 304 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1287;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGAGAC 25  
|||||  
Db 1013 TAAATGGTGGCGGTGGGGAGAC 989

RESULT 2  
AF181722  
LOCUS AF181722 Homo sapiens RU2AS (RU2) mRNA linear PRI 10-JAN-2000  
DEFINITION Homo sapiens RU2AS (RU2) mRNA, complete cds.  
ACCESSION AF181722  
VERSION AF181722.1 GI:6684531  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1382)  
AUTHORS Van Den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L., Devuyt,O., Lorge,F., Weynants,P. and Boon,T.  
TITLE A new antigen recognized by cytolytic T lymphocytes on a human kidney tumor results from reverse strand transcription

J. Exp. Med. 190 (12), 1793-1800 (1999)  
20069887  
10601354  
REFERENCE 2 (bases 1 to 1382)  
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilotte,L.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
Avenue Hippocrate, 74, Brussels 1200, Belgium

FEATURES  
source  
1. .1382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p22.1"  
1. .1382  
/gene="RU2"  
738. .992  
/gene="RU2"  
/codon\_start=1  
/product="RU2AS"  
/protein\_id="AAF23613.1"  
/db\_xref="GI:6684532"  
/translation="MDDDAAPRVEGVPVAVHKKHALHDGLRQVAGPAAAAHLPRWPPP  
QIARREAPPLSQRPHTQAGSPETNEKLTNPQVKEK"

BASE COUNT 355 a 373 c 344 g 310 t

ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1382;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGAGAC 25  
|||||  
Db 523 TAAATGGTGGCGGTGGGGAGAC 547

RESULT 3  
AF181720/c  
LOCUS AF181720 Homo sapiens RU2AS (RU2) gene, complete cds; and RU2S (RU2) gene,  
DEFINITION partial cds.  
ACCESSION AF181720  
VERSION AF181720.1 GI:6684526  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 4377)  
AUTHORS Van Den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L., Devuyt,O., Lorge,F., Weynants,P. and Boon,T.  
TITLE A new antigen recognized by cytolytic T lymphocytes on a human kidney tumor results from reverse strand transcription

J. Exp. Med. 190 (12), 1793-1800 (1999)  
20069887  
10601354  
REFERENCE 2 (bases 1 to 4377)  
AUTHORS Van den Eynde,B.J., Gaugler,B. and Pilotte,L.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,  
Avenue Hippocrate, 74, Brussels 1200, Belgium

FEATURES  
source  
1. .4377  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p22.1"  
1754. .>2348  
/gene="RU2"  
1754. .>2348  
/gene="RU2"  
/product="RU2S"  
complement(<1912. .>2166)

gene  
mRNA  
mRNA

CDS  
 /product="RU2AS"  
 complement(1912..2166)  
 /codon\_start=1  
 /product="RU2AS"  
 /protein\_id="AAF23611.1"  
 /db\_xref="GI:6684528"  
 /translation="MDDDAAPRVEGVVAVYKHALHDGLRQVAGPGAAAHLPWP  
 QLAASREAPPLSORPHRTQAGSPPETNEKLTNPQVKR"  
 2056..>2348  
 /gene="RU2"  
 /genomic\_start=1  
 /codon\_start=1  
 /product="RU2S"  
 /protein\_id="AAF23610.1"  
 /db\_xref="GI:6684527"  
 /translation="MSGSSARSSHLSPVVKSVLYRNGDPFVAGRRVVIHEKKVSF  
 EVLKEVTGGVQAPFGAVRNIYPRGHRIRKLDIQSGNYVAGQGEAFKKL"

BASE COUNT 1296 a 858 c 828 g 1380 t 15 others  
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 4377;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TAAATGGTGGCGGTGGGGAGAC 25  
 Db 2381 TAAATGGTGGCGGTGGGGAGAC 2357

RESULT 4  
 HSJ282H10 120029 bp DNA linear HTG 10-JUL-2001  
 LOCUS Homo sapiens chromosome 6 clone RP1-282H10 map p22.1-22.3, \*\*\*  
 DEFINITION SEQUENCING IN PROGRESS \*\*\*, 5 unordered pieces.  
 ACCESSION AL132672  
 VERSION AL132672.14 GI:14348905  
 KEYWORDS HTG; HTGS\_PRAISE1; HTGS\_CANCELLED.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 120029)

Direct Submission  
 Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Jun 12, 2001 this sequence version replaced gi:12331282.  
 ----- Genome Center

Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: dj282H10

----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: M13; W77815; 0% of reads  
 Chemistry: Dye-terminator ABI; 1% of reads  
 Chemistry: Dye-terminator Big Dye; 98% of reads  
 Consensus quality: 118224 bases at least Q40  
 Consensus quality: 118449 bases at least Q30  
 Consensus quality: 118634 bases at least Q20  
 Insert size: 119629; sum-of-contigs  
 Insert size: 117431; 9.3% error; agarose-fp  
 Quality coverage: 18.41x in Q20 bases; sum-of-contigs Quality  
 coverage: 19.36x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 23429: contig of 23429 bp in length  
 \* 23430 23529: gap of 100 bp  
 \* 23530 63600: contig of 40071 bp in length  
 \* 63601 63700: gap of 100 bp  
 \* 63701 100099: contig of 36399 bp in length  
 \* 100100 100199: gap of 100 bp  
 \* 100200 102366: contig of 2167 bp in length  
 \* 102367 102466: gap of 100 bp  
 \* 102467 120029: contig of 17563 bp in length.

#### FEATURES

source  
 1..120029  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="p22.1-22.3"  
 /clone="RP1-282H10"  
 /clone\_lib="RPC1-1"

misc\_feature  
 1..23429  
 /note="assembly\_fragment:05837  
 clone\_end:SP6  
 vector\_side:left"

misc\_feature  
 23530..63600  
 /note="assembly\_fragment:00423  
 fragment\_chain:1"

misc\_feature  
 63701..100099  
 /note="assembly\_fragment:02685  
 fragment\_chain:1"

misc\_feature  
 100200..102366  
 /note="assembly\_fragment:01789"

misc\_feature  
 102467..120029  
 /note="assembly\_fragment:02341  
 clone\_end:T7  
 vector\_side:right"

BASE COUNT 36483 a 24840 c 24637 g 33666 t 403 others  
 ORIGIN

Query Match 100.0%; Score 25; DB 2; Length 120029;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TAAATGGTGGCGGTGGGGAGAC 25

Db 21570 TAAATGGTGGCGGTGGGGAGAC 21594

RESULT 5  
 AL359713

LOCUS Human DNA sequence from clone RP11-95P3 on chromosome 6, complete  
 DEFINITION sequence.

ACCESSION AL359713

VERSION AL359713.25 GI:13938809

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 152966)

Kimberley, A.

Direct Submission

Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On May 3, 2001 this sequence version replaced gi:13446455.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr6>

RP11-95P3 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-95P3. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-95P3 is at 1 in this sequence. The true left end of clone RP1-73M23 is at 152867 in this sequence. The true right end of clone RP11-40E20 is at 17700 in this sequence.

## FEATURES

```

source
    1..152966
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="6"
    /clone="RP11-95P3"
    /clone_lib="RPCI-11.1"
    19..544
    /note="match: GSS: Em:AZ517849"
    24..379
    /note="match: GSS: Em:B63526"
    32..494
    /note="match: GSS: Em:AQ315706"
    70..589
    /note="match: GSS: Em:AQ285677"
    179..362
    /note="MIR repeat: matches 65..262 of consensus"
    521..582
    /note="L2 repeat: matches 2683..2741 of consensus"
    720..1022
    /note="AluX repeat: matches 1..303 of consensus"
    1327..1448
    /note="AluSg repeat: matches 181..298 of consensus"
    2264..2575
    /note="AluSg repeat: matches 1..311 of consensus"
    3014..3296
    /note="match: STS: Em:G27290"
    3014..3093
    /note="40 copies 2 mer ga 75% conserved"
    5024..5477
    /note="L1M1 repeat: matches 1012..1598 of consensus"
    5581..5859
    /note="AluSg repeat: matches 18..294 of consensus"
    5986..6359
    /note="L1MEC repeat: matches 2228..2259 of consensus"
    6367..6515
    /note="FIAM_C repeat: matches 1..143 of consensus"
    6971..7198
    /note="L2 repeat: matches 1512..1735 of consensus"
    8403..8838
    /note="LTR7 repeat: matches 1..445 of consensus"
    8799..9441
    /note="match: GSS: Em:B54691"
    complement(9393..9853)
    /note="match: GSS: Em:AQ285817"
    10311..10437
    /note="L2 repeat: matches 2612..2747 of consensus"

repeat_region
    10710..11022
    /note="AluJo repeat: matches 1..309 of consensus"
    12006..12156
    /note="L2 repeat: matches 2594..2748 of consensus"
    12182..13011
    /note="L2 repeat: matches 1374..2344 of consensus"
    13404..13631
    /note="114 copies 2 mer at 55% conserved"
    13479..13630
    /note="38 copies 4 mer tata 60% conserved"
    13785..14820
    /note="L1ME repeat: matches 4766..5798 of consensus"
    14928..16008
    /note="L1PB3 repeat: matches 5106..6149 of consensus"
    16029..16110
    /note="L1PB3 repeat: matches 5054..5133 of consensus"
    16428..16829
    /note="L1MC4 repeat: matches 6189..6603 of consensus"
    16868..17150
    /note="AluX repeat: matches 1..285 of consensus"
    17246..17991
    /note="L1MC4 repeat: matches 6608..7343 of consensus"
    17992..18675
    /note="MER67C repeat: matches 1..710 of consensus"
    18676..18729
    /note="L1MC4 repeat: matches 7343..7396 of consensus"
    18731..19458
    /note="MER4D repeat: matches 232..973 of consensus"
    complement(18929..19542)
    /note="match: GSS: Em:AQ342961"
    complement(19031..19542)
    /note="match: GSS: Em:AQ053552"
    complement(19065..19495)
    /note="match: GSS: Em:AQ545892"
    complement(19130..19546)
    /note="match: GSS: Em:B94529"
    19265..19495
    /note="match: GSS: Em:AQ373351 Em:AQ375803"
    match: STS: Em:G59266"
    complement(19269..19546)
    /note="match: GSS: Em:AQ663911"
    complement(19325..19546)
    /note="match: GSS: Em:AQ021494"
    complement(19337..19501)
    /note="match: GSS: Em:AQ268095"
    complement(19370..19546)
    /note="match: GSS: Em:AQ128461"
    19566..19739
    /note="AluSg/x repeat: matches 126..295 of consensus"
    19740..20100
    /note="L1MD3 repeat: matches 7391..7739 of consensus"
    21121..21429
    /note="AluJo repeat: matches 1..310 of consensus"
    22589..22728
    /note="70 copies 2 mer aa 60% conserved"
    22591..22670
    /note="20 copies 4 mer aaag 80% conserved"
    22686..22761
    /note="19 copies 4 mer aagg 93% conserved"
    23098..23393
    /note="AluJb repeat: matches 4..298 of consensus"
    23468..23578
    /note="L2 repeat: matches 2580..2700 of consensus"
    23980..24292
    /note="AluSg repeat: matches 2..310 of consensus"
    complement(27208..27677)
    /note="match: GSS: Em:AQ0702871"
    27819..27957
    /note="AluJb repeat: matches 163..299 of consensus"
    28701..28930
    /note="MIRIG repeat: matches 32..301 of consensus"
    29080..29171
    /note="MLT1H repeat: matches 433..526 of consensus"

```

```

repeat_region 29303..29711
/note="L2 repeat: matches 2323. .2710 of consensus"
repeat_region 29712..29834
/note="MTIG repeat: matches 29. .147 of consensus"
repeat_region 30140..30547
/note="L2 repeat: matches 1806. .2252 of consensus"
misc_feature complement(30720..31207)
/note="match: STS: Em:HS1J17"
repeat_region 30919..31120
/note="MER58A repeat: matches 1. .208 of consensus"
repeat_region 32109..32472
/note="THEIC repeat: matches 1. .371 of consensus"
repeat_region 33360..33511
/note="MIR repeat: matches 98. .250 of consensus"
repeat_region 34031..34331
/note="AluX repeat: matches 1. .297 of consensus"
repeat_region 34621..34732
/note="MIR repeat: matches 120. .232 of consensus"
repeat_region 34888..35007
/note="L2 repeat: matches 1836. .1959 of consensus"
repeat_region 35164..35461
/note="AluSg repeat: matches 2. .298 of consensus"
repeat_region 35532..35990
/note="L2 repeat: matches 2254. .2710 of consensus"
repeat_region 36631..36790
/note="L2 repeat: matches 2342. .2501 of consensus"
repeat_region 37251..37544
/note="AluX repeat: matches 1. .294 of consensus"
repeat_region 38446..38493
/note="L2 copies 4 mer caca 75% conserved"
repeat_region 38467..38492

Query Match 100.0%; Score 25; DB 9; Length 152966;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGTGGCGGTGGGGGAGAC 25
|||||
Db 53944 TAAATGGTGGCGGTGGGGGAGAC 53968

RESULT 6
AC115426/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-185A19, *** SEQUENCING IN PROGRESS
*** 52 unordered pieces.
ACCESSION AC115426
VERSION AC115426.2 GI:21737279
KEYWORDS HTG; HTGS-PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 98724)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haviak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,

```

```

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwar,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojhas,A., Rojuben,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 98724)
Worley,K.C.
Direct Submission
Submitted (19-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 98724)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:19549134.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOVP
Center clone name: CH230-185A19
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 42969 bases at least Q40
Consensus quality: 45549 bases at least Q30
Consensus quality: 47683 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 52 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be closed with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1001: contig of 1001 bp in length
* 1002 1101: gap of unknown length
* 1102 2113: contig of 1012 bp in length
* 2114 2213: gap of unknown length
* 2214 3557: contig of 1344 bp in length
* 3558 3657: gap of unknown length
* 3658 5062: contig of 1405 bp in length
* 5063 5162: gap of unknown length
* 5163 6344: contig of 1182 bp in length
* 6345 6444: gap of unknown length

```

\* 6445 7700: contig of 1256 bp in length  
\* 7701 7800: gap of unknown length  
\* 7801 8879: contig of 1079 bp in length  
\* 8880 8979: gap of unknown length  
\* 8980 10599: contig of 1620 bp in length  
\* 10600 10699: gap of unknown length  
\* 10700 12188: contig of 1489 bp in length  
\* 12189 12288: gap of unknown length  
\* 12289 13417: contig of 1129 bp in length  
\* 13418 13517: gap of unknown length  
\* 13518 14780: contig of 1263 bp in length  
\* 14781 14880: gap of unknown length  
\* 14881 15948: contig of 1068 bp in length  
\* 15949 16048: gap of unknown length  
\* 16049 17156: contig of 1108 bp in length  
\* 17157 17256: gap of unknown length  
\* 17257 18436: contig of 1170 bp in length  
\* 18437 18526: gap of unknown length  
\* 18527 19567: contig of 1041 bp in length  
\* 19568 19667: gap of unknown length  
\* 19668 20902: contig of 1235 bp in length  
\* 20903 21002: gap of unknown length  
\* 21003 22503: contig of 1501 bp in length  
\* 22504 23840: gap of unknown length  
\* 23841 23940: gap of unknown length  
\* 23941 25608: contig of 1668 bp in length  
\* 25609 25708: gap of unknown length  
\* 25709 27121: contig of 1413 bp in length  
\* 27122 27221: gap of unknown length  
\* 27222 28815: contig of 1594 bp in length  
\* 28816 28915: gap of unknown length  
\* 28916 30261: contig of 1346 bp in length  
\* 30262 30361: gap of unknown length  
\* 30362 32113: contig of 1752 bp in length  
\* 32114 32213: gap of unknown length  
\* 32214 33765: contig of 1552 bp in length  
\* 33766 33865: gap of unknown length  
\* 33866 35271: contig of 1406 bp in length  
\* 35272 35371: gap of unknown length  
\* 35372 36681: contig of 1310 bp in length  
\* 36682 36781: gap of unknown length  
\* 36782 38466: contig of 1685 bp in length  
\* 38467 38566: gap of unknown length  
\* 38567 40066: contig of 1500 bp in length  
\* 40067 40166: gap of unknown length  
\* 40167 41623: contig of 1457 bp in length  
\* 41624 41723: gap of unknown length  
\* 41724 43639: contig of 1916 bp in length  
\* 43640 43739: gap of unknown length  
\* 43740 45987: contig of 2248 bp in length  
\* 45988 46087: gap of unknown length  
\* 46088 48090: contig of 2003 bp in length  
\* 48091 48190: gap of unknown length  
\* 48191 49859: contig of 1669 bp in length  
\* 49860 49959: gap of unknown length  
\* 49960 52243: contig of 2284 bp in length  
\* 52244 52343: gap of unknown length  
\* 52344 53931: contig of 1608 bp in length  
\* 53932 54051: gap of unknown length  
\* 54052 55867: contig of 1816 bp in length  
\* 55868 55967: gap of unknown length  
\* 55968 57699: contig of 1732 bp in length  
\* 57700 57799: gap of unknown length  
\* 57800 59785: contig of 1966 bp in length  
\* 59786 59865: gap of unknown length  
\* 59866 61448: contig of 1583 bp in length  
\* 61449 61548: gap of unknown length  
\* 61549 63060: contig of 1512 bp in length  
\* 63061 63160: gap of unknown length  
\* 63161 64933: contig of 1793 bp in length  
\* 64934 65053: gap of unknown length  
\* 65054 68177: contig of 3124 bp in length

\* 68178 68277: gap of unknown length  
\* 68278 70398: contig of 2121 bp in length  
\* 70399 70498: gap of unknown length  
\* 70499 73152: contig of 2654 bp in length  
\* 73153 73252: gap of unknown length  
\* 73253 75978: contig of 2726 bp in length  
\* 75979 76078: gap of unknown length  
\* 76079 78390: contig of 2312 bp in length  
\* 78391 78490: gap of unknown length  
\* 78491 81417: contig of 2927 bp in length  
\* 81418 81517: gap of unknown length  
\* 81518 84420: contig of 2903 bp in length  
\* 84421 84520: gap of unknown length  
\* 84521 88161: contig of 3641 bp in length  
\* 88162 88261: gap of unknown length  
\* 88262 90893: contig of 2632 bp in length  
\* 90894 90993: gap of unknown length  
\* 90994 94552: contig of 3559 bp in length  
\* 94553 94652: gap of unknown length  
\* 94653 98724: contig of 4072 bp in length.

FEATURES  
source  
1..98724  
Location/Qualifiers  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"

Query Match 83.2%; Score 20.8; DB 2; Length 98724;  
Best Local Similarity 91.7%; Pred. No. 1.9e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATGGTGGCGGTGGGGAGAC 25  
|||||  
Db 13303 AAATGGTGGCGGGGGAGAC 13280

RESULT 7  
AL157396  
LOCUS  
DEFINITION Human DNA sequence from clone RP11-437J2 on chromosome 10, complete sequence.  
ACCESSION AL157396  
VERSION AL157396.9 GI:14018250  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 128796)  
BROWN, A.  
Direct Submission  
Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On May 11, 2001 this sequence version replaced gi:13751283.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations. Together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10



Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
 RP11-437J2 is from the library RPCI-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.choi.org/bacpac/home.htm>  
 VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone  
 RP11-437J2. It may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true right end of clone RP11-437J2 is at 128796 in this  
 sequence. The true right end of clone RP11-30E16 is at 100 in this  
 sequence.

FEATURES  
 source Location/Qualifiers  
 1..128796  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="10"  
 /clone="RP11-437J2"  
 /clone\_lib="RPCI-11.2"  
 BASE COUNT 40665 a 25975 c 24235 g 37921 t  
 ORIGIN

Query Match 83.2%; Score 20.8; DB 9; Length 128796;  
 Best Local Similarity 91.7%; Pred. No. 1.8e+02;  
 Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AATGGGTGGCGGTGGGGAGAC 25

||||||| | |||||

Db 18347 AATGGGTGGCGGTGGGGAGAC 18370

RESULT 8  
 AC091543  
 LOCUS AC091543 164773 bp DNA linear HTG 04-MAY-2001  
 DEFINITION Felis catus clone RP86-588L5, WORKING DRAFT SEQUENCE, 12 unordered  
 pieces.

ACCESSION AC091543.1 GI:13940632

VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Felis catus.

ORGANISM

REFERENCE 1 (bases 1 to 164773)  
 Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

AUTHORS Avele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,  
 Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S.,  
 Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E., Lee-Lin, S.-Q.,  
 Legaspi, R., Lin, M., Maduro, Q.L., Maduro, V.B., Masiello, C.,  
 Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,  
 Shevchenko, Y., Snyder, B., Stantripod, S., Thomas, J.W., Thomas, P.J.,  
 Tlonsong, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
 Wetherby, K.D., Zhang, L.-H. and Green, E.D.

Unpublished NISC Comparative Sequencing Initiative

2 (bases 1 to 164773)

Green, E.D.

Direct Submission

Submitted (04-MAY-2001) NIH Intramural Sequencing Center, 8717

Groveton Circle, Gaithersburg, MD 20877, USA

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: [nisc.mouse@nih.gov](mailto:nisc.mouse@nih.gov)

----- Project Information

Center project name: awf

Center clone name: 588L05

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 158805 bases at least Q40

Consensus quality: 160179 bases at least Q30

Consensus quality: 160900 bases at least Q20  
 Insert size: 175000; agarose-fp  
 Insert size: 163673; sum-of-contigs  
 Quality coverage: 10.73x in Q20 bases; agarose-fp  
 Quality coverage: 11.47x in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 12 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2359: contig of 2359 bp in length  
 \* 2360 2459: gap of unknown length  
 \* 2460 6460: contig of 4001 bp in length  
 \* 6461 6560: gap of unknown length  
 \* 6561 10554: contig of 3994 bp in length  
 \* 10555 10654: gap of unknown length  
 \* 10655 16088: contig of 5414 bp in length  
 \* 16089 16168: gap of unknown length  
 \* 16169 23033: contig of 6865 bp in length  
 \* 23034 23133: gap of unknown length  
 \* 23134 30910: contig of 7777 bp in length  
 \* 30911 31010: gap of unknown length  
 \* 31011 42000: contig of 10990 bp in length  
 \* 42001 42100: gap of unknown length  
 \* 42101 57878: contig of 15778 bp in length  
 \* 57879 57979: gap of unknown length  
 \* 57979 81265: contig of 23287 bp in length  
 \* 81266 81366: gap of unknown length  
 \* 81366 101649: contig of 20284 bp in length  
 \* 101650 101749: gap of unknown length  
 \* 101750 128418: contig of 26669 bp in length  
 \* 128419 128518: gap of unknown length  
 \* 128519 164773: contig of 36255 bp in length.

#### FEATURES

source

1..164773  
 /organism="Felis catus"  
 /db\_xref="taxon:9685"  
 /clone="RP86-588L5"  
 /clone\_lib="RP86"

1..2359

/note="assembly\_fragment"

2460..6460

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:right

6561..10554

/note="assembly\_fragment"

10655..16068

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right

16169..23033

/note="assembly\_fragment"

23134..30910

/note="assembly\_fragment"

31011..42000

/note="assembly\_fragment"

42101..57878

/note="assembly\_fragment"

57979..81265

/note="assembly\_fragment"

81366..101649

/note="assembly\_fragment"

101750..128418

/note="assembly\_fragment"

128519..164773

/note="assembly\_fragment"

BASE COUNT 37993 a 45086 c 43608 g 36964 t 1122 others  
 ORIGIN

## AUTHORS

TITLE  
JOURNAL

## COMMENT

Worley, K.C.  
Direct Submission  
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:20066113.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GUT1  
Center clone name: CH230-177D13  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 103705 bases at least Q40  
Consensus quality: 109417 bases at least Q30  
Consensus quality: 113895 bases at least Q20  
-----

Query Match 83.2%; Score 20.8; DB 2; Length 164773;  
Best Local Similarity 91.7%; Pred. No. 1.7e+02;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAAATGGGTGGCGGTGGGGGAGA 24

Db 36014 TAAATGGGTGGCGGTGGGGGAGA 36037

## RESULT 9

AC117062

LOCUS

DEFINITION

AC117062

AC117062.2

HTG: HTGS-PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 168203)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,

Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,

Barbala, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,

Buahy, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,

Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

Gorelli, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,

Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,

Homs, F., Howard, S., Huber, J., Huiyk, S., Hume, J., Jackson, L.E.,

Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,

Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,

Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,

Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,

Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,

Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,

Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,

Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,

Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,

Rivers, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G.,

Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,

Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,

Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,

Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,

Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,

Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,

Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,

Winstock, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 168203)

Worley, K.C.

Direct Submission

Submitted (06-APR-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 168203)

REFERENCE

## TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

3 (bases 1 to 168203)

\* 32397 32496: gap of unknown length  
\* 32497 33882: contig of 1386 bp in length  
\* 33882 33982: gap of unknown length  
\* 33982 33983: contig of 1416 bp in length  
\* 33983 33989: gap of unknown length  
\* 33989 34246: contig of 1748 bp in length  
\* 34246 37347: gap of unknown length  
\* 37347 37347: contig of 1796 bp in length  
\* 37347 39142: contig of 1796 bp in length  
\* 39142 39242: gap of unknown length  
\* 39242 41907: contig of 2664 bp in length  
\* 41907 42006: gap of unknown length  
\* 42006 43861: contig of 1855 bp in length  
\* 43861 43962: gap of unknown length  
\* 43962 45577: contig of 1616 bp in length  
\* 45577 45678: gap of unknown length  
\* 45678 47112: contig of 1435 bp in length  
\* 47112 47212: gap of unknown length  
\* 47212 48684: contig of 1472 bp in length  
\* 48684 48784: gap of unknown length  
\* 48784 51006: contig of 2222 bp in length  
\* 51006 51106: gap of unknown length  
\* 51106 53696: contig of 2590 bp in length  
\* 53696 53797: gap of unknown length  
\* 53797 55270: contig of 1474 bp in length  
\* 55270 55371: gap of unknown length  
\* 55371 56901: contig of 1531 bp in length  
\* 56901 57001: gap of unknown length  
\* 57001 59230: contig of 2229 bp in length  
\* 59230 59330: gap of unknown length  
\* 59330 61357: contig of 2027 bp in length  
\* 61357 61457: gap of unknown length  
\* 61457 63549: contig of 2092 bp in length  
\* 63549 63649: gap of unknown length  
\* 63649 68661: contig of 3212 bp in length  
\* 68661 68961: gap of unknown length  
\* 68961 69191: contig of 2230 bp in length  
\* 69191 69292: contig of 1841 bp in length  
\* 69292 71132: contig of 1841 bp in length  
\* 71132 73610: contig of 2378 bp in length  
\* 73610 73710: gap of unknown length  
\* 73710 76132: contig of 2422 bp in length  
\* 76132 76232: gap of unknown length  
\* 76232 78750: contig of 2518 bp in length  
\* 78750 78850: gap of unknown length  
\* 78850 82498: contig of 3648 bp in length  
\* 82498 82598: gap of unknown length  
\* 82598 85698: contig of 3100 bp in length  
\* 85698 85798: gap of unknown length  
\* 85798 88048: contig of 2250 bp in length  
\* 88048 88148: gap of unknown length  
\* 88148 92440: contig of 4292 bp in length  
\* 92440 92540: gap of unknown length  
\* 92540 96697: contig of 4157 bp in length  
\* 96697 96797: gap of unknown length  
\* 96797 102170: contig of 5373 bp in length  
\* 102170 102270: gap of unknown length  
\* 102270 106111: contig of 3841 bp in length  
\* 106111 106211: gap of unknown length  
\* 106211 110828: contig of 4617 bp in length  
\* 110828 110928: gap of unknown length  
\* 110928 116356: contig of 5428 bp in length  
\* 116356 116357: gap of unknown length  
\* 116357 121945: contig of 5489 bp in length  
\* 121945 122046: gap of unknown length  
\* 122046 128572: contig of 6527 bp in length

Query Match 81.6%; Score 20.4; DB 2; Length 168203;  
Best Local Similarity 95.5%; Pred. No. 2.4e+02;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 AAATGGTGGGGTGGGGGAG 23  
||||||| |||||

Db 26252 AAATGGTGGGGTGGGGGAG 26273

RESULT 10  
TAHYNAS  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

TAHYNAS  
Tahyna virus S segment RNA (strain 92).  
268497  
268497.1 GI:1200157  
NS protein; S segment.  
Tahyna virus.  
Tahyna virus  
Viruses; ssRNA negative-strand viruses; Bunyaviridae;  
Orthobunyavirus.  
1 (bases 1 to 977)  
Plyusnin,A.  
Direct Submission  
Submitted (03-JAN-1996) Alexander Plyusnin, Department of Virology,  
Haartman Institute, Helsinki University, Haartmaninkatu 3,  
Helsinki, FIN-00014, Finland  
2 (bases 1 to 977)  
Vapalahti,O., Plyusnin,A., Cheng,Y., Manni,T.,  
Brummer-Korvenkontio,M. and Vaheri,A.  
Inkoo and Tahyna, the European California serogroup bunyaviruses:  
sequence and phylogeny of the S RNA segment  
J. Gen. Virol. 77 (pt 8), 1769-1774 (1996)  
96332514  
8760424  
On Feb 23, 1996 this sequence version replaced gi:1164992.  
Location/Qualifiers  
1..977  
/organism="Tahyna virus"  
/strain="92"  
/db\_xref="taxon:45270"  
/chromosome="S-segment"  
79..786  
/codon\_start=1  
/product="N protein"  
/protein\_id="CAA92806.1"  
/db\_xref="GI:1164993"  
/translation="MSDLVFDVASTANGFDPDAGYVDFCIKHGEAINLHVSRIFFL  
NAAKAALARKPERKAPKFGQVENVYHFGNRRNPIDNNDLTHRLYGLARV  
VLEQFKENEDAAQRELKTTVINPTAESINRWNGAEIYLAFFPGTMEFTFNFP  
LTIGIYRVKGMDDPYLLKALRQYGSUTADKWSQRTTAIAKSLKDVQLKWRGG  
LSDTARFLQFGRLP"  
98..391  
/codon\_start=1  
/product="NS protein"  
/protein\_id="CAA92806.1"  
/db\_xref="GI:1164994"  
/db\_xref="SPTREMBL:Q88472"  
/translation="MMSHPVQMDLILMQGMWTSVLNMGKQLISIPLGSSSLMPQKPK  
LLSLVSRGRGLVNLNLSGRWRSIIIFLETGTQLITITLPSTGCTGIWLDGC"

BASE COUNT 305 a 180 c 233 g 259 t  
ORIGIN

Query Match 80.8%; Score 20.2; DB 14; Length 977;  
Best Local Similarity 88.0%; Pred. No. 8.8e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 TAAATGGTGGGGTGGGGGAGAC 25  
||||||| ||||| ||||| ||  
Db 894 TAAATGGTGGGTGGTAGGGGAC 918

RESULT 11  
AC119906/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS

AC119906  
Mus musculus clone RP24-239D13, LOW-PASS SEQUENCE SAMPLING.  
AC119906  
AC119906.1 GI:20389564  
HTG; HTGS\_PHASE0.

```

SOURCE      Mus musculus.
ORGANISM    Mus musculus
REFERENCE   Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS     Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
TITLE       1 (bases 1 to 70905)
JOURNAL     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE   Mus musculus, clone RP24-239D13
AUTHORS     Unpublished
TITLE       2 (bases 1 to 70905)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazarro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
            Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
TITLE       Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
JOURNAL     Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L25284
            Center clone name: 239_D_13
            -----
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 691: contig of 691 bp in length
            * 792 791: gap of 100 bp
            * 1481 1481: contig of 690 bp in length
            * 1582 2295: contig of 714 bp in length
            * 2296 2395: gap of 100 bp
            * 2396 3014: contig of 619 bp in length
            * 3015 3114: gap of 100 bp
            * 3115 3850: contig of 736 bp in length
            * 3851 3950: gap of 100 bp
            * 3951 4619: contig of 669 bp in length
            * 4620 4719: gap of 100 bp
            * 4720 5407: contig of 688 bp in length
            * 5408 5507: gap of 100 bp
            * 5508 6196: contig of 689 bp in length
            * 6197 6296: gap of 100 bp
            *
            * 6297 7009: contig of 713 bp in length
            * 7010 7109: gap of 100 bp
            * 7110 7805: contig of 696 bp in length
            * 7806 7905: gap of 100 bp
            * 7906 8615: contig of 710 bp in length
            * 8616 8715: gap of 100 bp
            * 8716 9429: contig of 714 bp in length
            * 9430 9529: gap of 100 bp
            * 9530 10249: contig of 720 bp in length
            * 10250 10349: gap of 100 bp
            * 10350 11059: contig of 710 bp in length
            * 11060 11159: gap of 100 bp
            * 11160 11856: contig of 697 bp in length
            * 11857 11956: gap of 100 bp
            * 11957 12632: contig of 676 bp in length
            * 12633 12732: gap of 100 bp
            * 12733 13425: contig of 693 bp in length
            * 13426 13525: gap of 100 bp
            * 13526 14193: contig of 668 bp in length
            * 14194 14293: gap of 100 bp
            * 14294 14959: contig of 666 bp in length
            * 14960 15059: gap of 100 bp
            * 15060 15752: contig of 693 bp in length
            * 15753 15852: gap of 100 bp
            * 15853 16545: contig of 693 bp in length
            * 16546 16645: gap of 100 bp
            * 16646 17336: contig of 691 bp in length
            * 17337 17436: gap of 100 bp
            * 17437 18129: contig of 693 bp in length
            * 18130 18229: gap of 100 bp
            * 18230 18944: contig of 715 bp in length
            * 18945 19044: gap of 100 bp
            * 19045 19763: contig of 719 bp in length
            * 19764 19863: gap of 100 bp
            * 19864 20586: contig of 723 bp in length
            * 20587 21421: contig of 735 bp in length
            * 21422 21521: gap of 100 bp
            * 21522 22211: contig of 690 bp in length
            * 22212 22311: gap of 100 bp
            * 22312 22962: contig of 651 bp in length
            * 22963 23062: gap of 100 bp
            * 23063 23748: contig of 686 bp in length
            * 23749 23848: gap of 100 bp
            * 23849 24535: contig of 687 bp in length
            * 24536 24635: gap of 100 bp
            * 24636 25329: contig of 694 bp in length
            * 25330 25429: gap of 100 bp
            * 25430 26129: contig of 700 bp in length
            * 26130 26229: gap of 100 bp
            * 26230 26929: contig of 700 bp in length
            * 26930 27029: gap of 100 bp
            * 27030 27732: contig of 703 bp in length
            * 27733 27832: gap of 100 bp
            * 27833 28548: contig of 716 bp in length
            * 28549 28648: gap of 100 bp
            * 28649 29367: contig of 719 bp in length
            * 29368 29467: gap of 100 bp
            * 29468 30168: contig of 701 bp in length
            * 30169 30268: gap of 100 bp
            * 30269 30961: contig of 693 bp in length
            * 30962 31061: gap of 100 bp
            * 31062 31759: contig of 698 bp in length
            * 31760 31859: gap of 100 bp
            * 31860 32555: contig of 696 bp in length
            * 32556 32655: gap of 100 bp
            * 32656 33321: contig of 666 bp in length
            * 33322 33421: gap of 100 bp
            * 33422 34116: contig of 695 bp in length
            * 34117 34216: gap of 100 bp
            * 34217 34907: contig of 691 bp in length
            * 34908 35007: gap of 100 bp
            * 35008 35706: contig of 699 bp in length

```



Db 97998 TTAGTGGTTGGCGGTGGGGAGAC 98022

# RESULT 13

## AC095226

### LOCUS

### DEFINITION

### AC095226

### VERSION

### AC095226.3

### KEYWORDS

### HTG: HTGS\_PHASE1

### SOURCE

### Norway rat.

### ORGANISM

### Rattus norvegicus

### Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

### Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

### Rattus

### REFERENCE

### 1 (bases 1 to 126441)

### Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,

### Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayale, M., Banks, T.,

### Barbaria, J., Benton, J., Blmage, K., Blankenburg, K., Bonnin, D.,

### Bouck, J., Bowles, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,

### Buhay, C., Burch, P., Burrett, C., Burrett, K.L., Byrd, N.C.,

### Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,

### Chen, G., Chen, K., Chen, Z., Chowdhry, I., Christopoulos, C.,

### Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,

### Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,

### Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

### Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

### Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,

### Falls, T., Ferraguto, D., Flegg, N., Ford, J., Foster, P., Frantz, P.,

### Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,

### Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,

### Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,

### Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,

### Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,

### Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,

### Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

### Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,

### Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H.,

### Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,

### Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,

### Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M.,

### Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,

### Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,

### Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,

### Oraquyne, N., Oviedo, R., Pace, A., Payton, B., Peary, J., Perez, L.,

### Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,

### Rivers, M., Rojas, A., Rojibokan, I., Roife, M., Ruiz, S., Savery, G.,

### Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,

### Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,

### Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,

### Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,

### Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,

### Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,

### Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,

### Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,

### Weinstock, G. and Gibbs, R.

### Direct Submission

### Unpublished

### 2 (bases 1 to 126441)

### Worley, K.C.

### Direct Submission

### Submitted (16-SEP-2001)

### Human Genome Sequencing Center, Department

### of Molecular and Human Genetics, Baylor College of Medicine, One

### Baylor Plaza, Houston, TX 77030, USA

### 3 (bases 1 to 126441)

### Worley, K.C.

### Direct Submission

### Submitted (11-JUL-2002)

### Human Genome Sequencing Center, Department

### of Molecular and Human Genetics, Baylor College of Medicine, One

### Baylor Plaza, Houston, TX 77030, USA

### On Jul 10, 2002 this sequence version replaced gi:17942462.

### ----- Genome Center

### COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDKL

Center clone name: CH230-9f16

----- Summary Statistics

Sequencing vector: plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 51063 bases at least Q40

Consensus quality: 56683 bases at least Q30

Consensus quality: 59972 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 72 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1001: contig of 1001 bp in length

\* 1002 1101: gap of unknown length

\* 1102 1114: contig of 1013 bp in length

\* 2115 2214: gap of unknown length

\* 2215 3221: contig of 1007 bp in length

\* 3222 3321: gap of unknown length

\* 3322 4870: contig of 1549 bp in length

\* 4871 4970: gap of unknown length

\* 4971 6027: contig of 1057 bp in length

\* 6028 6127: gap of unknown length

\* 6128 7164: contig of 1037 bp in length

\* 7165 7264: gap of unknown length

\* 7265 8718: contig of 1454 bp in length

\* 8719 9973: contig of 1155 bp in length

\* 9974 10073: gap of unknown length

\* 10074 11224: contig of 1251 bp in length

\* 11225 11424: gap of unknown length

\* 11425 12606: contig of 1182 bp in length

\* 12607 12706: gap of unknown length

\* 12707 13893: contig of 1187 bp in length

\* 13894 13993: gap of unknown length

\* 13994 15420: contig of 1427 bp in length

\* 15421 15520: gap of unknown length

\* 15521 16956: contig of 1436 bp in length

\* 16957 17056: gap of unknown length

\* 17057 18723: contig of 1667 bp in length

\* 18724 18823: gap of unknown length

\* 18824 20632: contig of 1809 bp in length

\* 20633 20732: gap of unknown length

\* 20733 22245: contig of 1513 bp in length

\* 22246 22345: gap of unknown length

\* 22346 23660: contig of 1315 bp in length

\* 23661 23760: gap of unknown length

\* 23761 25583: contig of 1823 bp in length

\* 25584 25683: gap of unknown length

\* 25684 26708: contig of 1025 bp in length

\* 26709 26808: gap of unknown length

\* 26809 28372: contig of 1564 bp in length

\* 28373 28472: gap of unknown length

\* 28473 29978: contig of 1506 bp in length

\* 29979 30078: gap of unknown length

\* 30079 31179: contig of 1101 bp in length

\* 31180 31279: gap of unknown length

\* 31280 32839: contig of 1560 bp in length

\* 32840 32939: gap of unknown length

\* 32940 34187: contig of 1248 bp in length

\* 34188 34287: gap of unknown length

\* 34288 35412: contig of 1125 bp in length  
\* 35112: gap of unknown length  
\* 35513 36750: contig of 1238 bp in length  
\* 36751 36850: gap of unknown length  
\* 36851 38069: contig of 1219 bp in length  
\* 38070 38169: gap of unknown length  
\* 38170 39478: contig of 1309 bp in length  
\* 39479 41384: gap of unknown length  
\* 39579 41384: contig of 1806 bp in length  
\* 41385 41484: gap of unknown length  
\* 41485 42713: contig of 1229 bp in length  
\* 42714 42813: gap of unknown length  
\* 42814 44935: contig of 2122 bp in length  
\* 44936 45035: gap of unknown length  
\* 45036 46222: contig of 1487 bp in length  
\* 46223 46222: gap of unknown length  
\* 46223 48205: contig of 1583 bp in length  
\* 48206 48305: gap of unknown length  
\* 48306 49613: contig of 1308 bp in length  
\* 49614 49713: gap of unknown length  
\* 49714 51293: contig of 1580 bp in length  
\* 51294 51393: gap of unknown length  
\* 51394 52698: contig of 1305 bp in length  
\* 52699 52798: gap of unknown length  
\* 52799 53988: contig of 1190 bp in length  
\* 53989 54088: gap of unknown length  
\* 54089 55758: contig of 1670 bp in length  
\* 55759 55858: gap of unknown length  
\* 55859 57497: contig of 1639 bp in length  
\* 57498 57597: gap of unknown length  
\* 57598 59450: contig of 1853 bp in length  
\* 59451 59550: gap of unknown length  
\* 59551 61106: contig of 1556 bp in length  
\* 61107 61206: gap of unknown length  
\* 61207 62965: contig of 1759 bp in length  
\* 62966 63065: gap of unknown length  
\* 63066 64703: contig of 1638 bp in length  
\* 64704 64803: gap of unknown length  
\* 64804 66663: contig of 1860 bp in length  
\* 66664 66763: gap of unknown length  
\* 66764 68181: contig of 1418 bp in length  
\* 68182 68281: gap of unknown length  
\* 68282 70094: contig of 1813 bp in length  
\* 70095 70194: gap of unknown length  
\* 70195 72255: contig of 2061 bp in length  
\* 72256 72355: gap of unknown length  
\* 72356 73703: contig of 1348 bp in length  
\* 73704 73803: gap of unknown length  
\* 73804 74834: contig of 1031 bp in length  
\* 74835 74934: gap of unknown length  
\* 74935 76180: contig of 1246 bp in length  
\* 76181 76280: gap of unknown length  
\* 76281 78118: contig of 1838 bp in length  
\* 78119 78218: gap of unknown length  
\* 78219 79980: contig of 1762 bp in length  
\* 79981 80080: gap of unknown length  
\* 80081 81696: contig of 1616 bp in length  
\* 81697 81796: gap of unknown length  
\* 81797 83379: contig of 1583 bp in length

Query Match 80.0%; Score 20; DB 2; Length 126441;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 GGGTGGCGGTGGGGAGAC 25  
|||||  
Db 16782 GGGTGGCGGTGGGGAGAC 16801

RESULT 14  
ARI29606/c 12141 bp DNA linear PAT 16-MAY-2001  
LOCUS ARI29606  
DEFINITION Sequence 10 from patent US 6187545.

ACCESSION ARI29606  
VERSION ARI29606.1 GI:14117503  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 12141)  
AUTHORS McKay, R., Butler, M.M., Wyatt, J. and Cowsett, L.M.  
TITLE Antisense modulation of pepck-cytosolic expression  
JOURNAL Patent: US 6187545-A 10 13-FEB-2001;  
FEATURES Location/Qualifiers  
source 1. 12141  
BASE COUNT 3122 a 2924 c 3087 g 3008 t  
ORIGIN  
Query Match 79.2%; Score 19.8; DB 6; Length 12141;  
Best Local Similarity 91.3%; Pred. No. 7.2e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TAAATGGTGGCGGTGGGGAG 23  
|||||  
Db 8237 TAAATGGAGGGAGGTGGGGAG 8215

RESULT 15  
AF009605/c 12141 bp DNA linear ROD 25-FEB-2000  
LOCUS Mus musculus phosphoenolpyruvate carboxykinase (PEPCK) gene,  
DEFINITION complete cds.  
ACCESSION AF009605  
VERSION AF009605.1 GI:4102181  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 12141)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 12141)  
AUTHORS Williams, C.P., Postic, C., Robin, D., Robin, P., Parrinello, J.,  
Shelton, K., Printz, R.L., Magnuson, M.A., Granner, D.K., Forest, C. and  
Chalkley, R.  
TITLE Isolation and characterization of the mouse cytosolic  
phosphoenolpyruvate carboxykinase (GTP) gene: evidence for  
tissue-specific hypersensitive sites  
JOURNAL Mol. Cell. Endocrinol. 148 (1-2), 67-77 (1999)  
MEDLINE 99236961  
PUBMED 10221772  
REFERENCE 2 (bases 1 to 12141)  
AUTHORS Williams, C.P., Postic, C., Parrinello, J., Shelton, K., Printz, R.L.,  
Granner, D.K., Magnuson, M.A. and Chalkley, R.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1997) Molecular Physiology and Biophysics,  
Vanderbilt University, 21st and Garland Avenues, Nashville, TN  
37232, USA  
FEATURES Location/Qualifiers  
source 1. 12141  
/organism="Mus musculus"  
/strain="129 SV"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="103.0"  
5495. 11764  
/gene="PEPCK"  
join(5495..5597,5854..6118,6440..6621,7206..7409,  
7682..7869,8444..8606,9418..9642,9735..9866,10502..10597,  
10701..11764)  
/gene="PEPCK"  
/product="phosphoenolpyruvate carboxykinase"  
5499. 5597  
/gene="PEPCK"  
/number=1  
5854. 6118  
/gene="PEPCK"

```

CDS
/number-2
Join(5895..6118,6440..6621,7206..7409,7682..7869,
8444..8606,9418..9642,9735..9866,10502..10597,
10701..11155)
/gene="PEPCK"
/codon_start=1
/product="phosphoenolpyruvate carboxykinase"
/protein_id="AAD01427.1"
/db_xref="GI:4102182"
/translation="MPOLHNLGLDPSAKVIOGSLDLPQAVRKFFVEGNAQLCQPEYIH
ICDSSEYQGQLAHMQBEGVIRKLKKYDNCWLAITDFDVARIESKVIITIQEORDT
VPIPTGUSQLGRWSEEDFEKAFNRPFGCMKRTMYVIFPSMGPLGSPKAKIGIEL
TDSPTVASMRTIMTGMISVLEALGDEFIKCLHSVGCPPLPKPLVNNWACNPELTL
IAHLDPREIISFGSGYGNLLGKCKFALRIASLAKKEGWLAEHMLILGITNPEGK
KYLAAAPPSAGCKTNLAMNPSLPWKVECVGDDIANMKDQGNLRAINPENGFFG
VAPGTSVKTNPNAIKTIQNTIFTNVAETSDGGVYEGIDEPLAPGVITTSKKNKEWR
PDABFCAPNSRFTCPASQCPIIDPAWESPEGVPIEGILFGRRRPEGVPLVYEALSW
OHGVFGAAMSEATAAAEHKGIIMHDPFAMRPFYNGFYGLAHWLSMAHRPAKLL
PKIFHVNFRKDKGLFMPGFGNSRVLEWMFRIEGEDSAKLTPIGYIPKENALNL
KGLGVNVEELFGISKFEWEVEIDRYLEDQVNTDLPYEIERELRALKQRISQM"
6440..6621
/exon
/gene="PEPCK"
/number-3
7206..7409
/gene="PEPCK"
/number-4
7682..7869
/gene="PEPCK"
/number-5
8444..8606
/gene="PEPCK"
/number-6
9418..9642
/gene="PEPCK"
/number-7
9735..9866
/gene="PEPCK"
/number-8
10502..10597
/gene="PEPCK"
/number-9
10701..111764
/gene="PEPCK"
/number-10
BASE COUNT 3122 a 3087 g 3008 t
ORIGIN
Query Match 79.2%; Score 19.8; DB 10; Length 12141;
Best Local Similarity 91.3%; Pred. No. 7.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAAATGGTGGCGGTGGGGGAG 23
||||||| ||| |||||
Db 8237 TAAATGGAGGAGGTGGGGGAG 8215
```

Search completed: January 19, 2003, 00:44:32  
Job time : 1868 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 18, 2003, 22:00:24 ; Search time 212.5 Seconds  
(without alignments)  
264.941 Million cell updates/sec

Title: US-09-674-593-9

Perfect score: 25

Sequence: 1 tagcgtgttgaaagggtagcaca 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAZ36648	PCR primer VDE120
2	25	100.0	1382	AAZ36643	Human tumour rejec
3	19.8	79.2	2181	AAT12500	Non-infective flick
4	19.8	79.2	2181	AAV59451	Plasmid construct
5	19.8	79.2	2418	AAH81162	Encodes Western su
6	19.8	79.2	11141	AAT44469	Tick-borne encephal
7	19.8	79.2	11141	AAT15820	TBE virus strain N
8	19.2	76.8	5098	AAD13425	Murine kit/stem ce
9	18.2	72.8	2055	AAV59954	DNA encoding preme

10	18.2	72.8	128139	24	AAI64291	RRV genome nucleot
11	18.2	72.8	133719	21	AAC64754	Macaca mulatta rha
12	18	72.0	372	24	ABN95473	Gene #1971 used to
13	18	72.0	372	24	ABL65082	Lung cancer relate
14	18	72.0	372	24	ABL66047	Lung cancer relate
15	18	72.0	372	24	ABL66942	Lung cancer relate
16	17.8	71.2	1464	23	AAV70831	DNA encoding novel
17	17.8	71.2	2695	22	AAI99567	Human expressed po
18	17.8	71.2	3685	24	AAV80777	Human DNA for pote
19	17.6	70.4	421	24	ABN73834	Bovine embryonic g
20	17.6	70.4	3690	22	AAV46183	Human DNA encoding
21	17.6	70.4	3902	24	AAD28148	Human secretion an
22	17.6	70.4	4949	21	AAC76989	Human ORFX ORF2544
23	17.2	68.8	434	22	AAV37291	Novel human diagno
24	17	68.0	66	16	AAT21811	Human gene signatu
25	17	68.0	361	21	AAC25567	Human secreted pro
26	17	68.0	578	22	ABA61347	Human foetal liver
27	17	68.0	578	22	AAK09646	Human brain expres
28	17	68.0	578	22	AAK35537	Human bone marrow
29	17	68.0	578	22	AAI41252	Probe #9938 used t
30	17	68.0	578	24	ABS09866	Human genome-deriv
31	17	68.0	1288	24	ABK35946	cDNA sequence #337
32	17	68.0	1309	16	AAQ94779	Regulator of neuro
33	17	68.0	1309	24	AAD37027	Mouse RNP-1 regula
34	17	68.0	5721	22	AAI04025	Human reproductive
35	17	68.0	5721	22	AAI04026	Human reproductive
36	17	68.0	16939	22	AAI04817	Human reproductive
37	17	68.0	16939	23	ABL97711	Human testicular a
38	17	68.0	22689	22	AAI04819	Human reproductive
39	17	68.0	22689	23	ABL97713	Human testicular a
40	17	68.0	40059	23	ABL15994	Drosophila melanog
41	16.8	67.2	49	18	AAT80484	Hepatoma AS-30D Ty
42	16.8	67.2	679	22	AAV32336	Human cDNA encodin
43	16.8	67.2	969	24	ABN66142	Streptococcus poly
44	16.8	67.2	5150	18	AAT78598	AS-30D tumour Type
45	16.6	66.4	158	24	ABA90249	Human ORF56 coding

#### ALIGNMENTS

##### RESULT 1

AAZ36648  
ID AAZ36648 standard; cDNA; 25 BP.

XX AAZ36648;

XX 22-FEB-2000 (first entry)

XX PCR primer VDE120 used to amplify RUR-1 antisense cDNA sequence.

XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW leukaemia; PCR primer; ss.

XX Synthetic.

OS Homo sapiens.

XX WO9958546-A1.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US10424.

XX 13-MAY-1998; 98US-0085318.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Van Den Eynde B, Boon-Falleur T;

XX WPI; 2000-053076/04.

XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,

```

PT e.g. treatment of cancers -
XX
XX Example 1; Page 44; 75pp; English.
XX
CC PCR primers AA226647-48 were used to amplify the antisense cDNA sequence
CC of human tumour rejection antigen RUR-1. The RUR-1 antisense sequence
CC is the antisense strand of a ubiquitously expressed gene. The antisense
CC strand codes for a polypeptide which is preferentially expressed in
CC tumour samples and tumour-derived cells lines. The polypeptide is
CC unrelated to any TRAP protein. The sequence was isolated from a renal
CC cell carcinoma line LB9211-RCC. The RUR-1 nucleic acids and polypeptides
CC can be used for diagnosis, prognosis or treatment of a disorder
CC characterized by the expression of a RUR-1 antisense cDNA molecule or
CC an expression product, such as cancers, e.g. renal cell carcinoma,
CC colorectal carcinoma, melanoma, sarcoma or leukaemia.
XX
XX Sequence 25 BP; 7 A; 3 C; 9 G; 6 T; 0 other;
XX
Query Match 100.0%; Score 25; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGCTGTTTGGAAAGGGTAGCACAC 25
Db 1 TAGGCTGTTTGGAAAGGGTAGCACAC 25
RESULT 2
AA236643/C
ID AA236643 standard; cDNA; 1382 BP.
XX
AC AA236643;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human tumour rejection antigen RUR-1 antisense cDNA sequence.
XX
KW Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 738..992
FT /*tag= a
FT primer_bind 523..547
FT /*tag= b
FT primer_bind /*note= "binding site for primer VDE119 (see AA236647)"
FT /*tag= c
FT /*note= "binding site for primer VDE120 (see AA236648)"
XX
XX WO9558546-A1.
XX
XX 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US10424.
XX
XX 13-MAY-1998; 98US-0085318.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
XX
XX Van Den Eynde B, Boon-Falleur T;
XX
XX WPI; 2000-053076/04.
XX
XX P-PSDB; AAY53809.
XX
XX New isolated tumour rejection antigen RUR-1 nucleic acids, used for,
XX e.g. treatment of cancers -
XX Claim 4; Fig 5; 75pp; English.
XX

```

---

```

CC The present sequence represents the antisense cDNA sequence of human
CC tumour rejection antigen RUR-1. The present sequence is the antisense
CC strand of a ubiquitously expressed gene. The antisense strand codes for
CC a polypeptide which is preferentially expressed in tumour samples and
CC tumour-derived cells lines. The polypeptide is unrelated to any TRAP
CC protein. The sequence was isolated from a renal cell carcinoma line
CC LB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for
CC diagnosis, prognosis or treatment of a disorder characterized by the
CC expression of a RUR-1 antisense cDNA molecule or an expression product,
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,
CC melanoma, sarcoma or leukaemia.
XX
XX Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 other;
XX
Query Match 100.0%; Score 25; DB 21; Length 1382;
Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGCTGTTTGGAAAGGGTAGCACAC 25
Db 1304 TAGGCTGTTTGGAAAGGGTAGCACAC 1280
RESULT 3
AAT12500
ID AAT12500 standard; DNA; 2181 BP.
XX
AC AAT12500;
XX
DT 30-AUG-1996 (first entry)
XX
DE Non-infective Tick-borne encephalitis virus construct pSV-P6Wt.
XX
KW Tick-borne encephalitis virus; TBE virus; vaccine; non-infective;
KW Flavivirus; membrane-associated protein E; prM; subviral particle;
KW ss.
XX
XX Tick-borne Encephalitis Virus.
XX
FH Key Location/Qualifiers
CDS 1..2166
FT /*tag= a
FT mat_peptide 94..585
FT /*tag= b
FT /*product= prM+M
FT /*note= "prM = 94..360; M = 361..585"
FT mat_peptide 586..2073
FT /*tag= c
FT /*product= membrane-associated_protein_E
FT mat_peptide 1..93
FT /*tag= d
FT /*product= C_(residues_86..116)
FT /*note= "encodes C-terminal region of core protein"
FT mat_peptide 2074..2163
FT /*tag= e
FT /*product= NS1_(residues_1..29)
FT /*note= "encodes N-terminal region of non-structural
FT protein NS1"
XX
XX EP691404-A2.
XX
XX 10-JAN-1996.
XX
XX 06-JUL-1995; 95EP-0890132.
XX
XX 23-MAY-1995; 95AT-0000871.
XX 08-JUL-1994; 94AT-0001352.
XX
XX (IMMO ) IMMUNO AG.
XX
XX Allison S, Heinz FX, Kunz C, Mandl C;
XX WPI; 1996-059736/07.
XX

```

DR P-PSDB; AAR88714.  
 XX Vaccine for protection against tick borne encephalitis virus -  
 PT comprises non-infectious subviral particles contg. at least the  
 PT complete E protein in native form, or nucleic acid encoding this  
 PT protein  
 XX  
 PS Example 1; Fig 3; 34pp; German.  
 XX  
 CC Parts of the TBE virus sequence coding for proteins E and prM+M were  
 CC amplified using PCR. Sequence analysis of the amplified fragments  
 CC showed that most contained mutations. One clone which encoded full-  
 CC length prM+M was combined with a second clone which encoded full-  
 CC length protein E in their wild-type forms. The combined coding  
 CC insert was incorporated into commercially available vector pSVbeta  
 CC from which the beta-galactosidase gene and part of the polylinker  
 CC had been removed. The resulting construct was designated pSV-Pewt  
 CC and was suitable for liposome-mediated transfection of COS-1 cells.  
 CC The cells were cultured to produce subviral particles for use in  
 CC a vaccine to protect mice against TBE virus infection. The present  
 CC sequence is that of the TBE-derived insert of plasmid pSV-Pewt.  
 XX  
 SQ Sequence 2181 BP; 558 A; 451 C; 686 G; 486 T; 0 other;  
 Query Match 79.2%; Score 19.8; DB 17; Length 2181;  
 Best Local Similarity 91.3%; Pred. No. 11;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TAGGCTGTTTGGAAAGGCTAGCA 23  
 Db 900 TGGACTGTTTGGAAAGGCTAGCA 922  
 RESULT 4  
 AAV59451  
 ID AAV59451 standard; DNA; 2181 BP.  
 XX  
 AC AAV59451;  
 XX  
 DT 15-JAN-1999 (first entry)  
 XX  
 DE Plasmid construct SV-Pewt DNA.  
 XX  
 KW Vaccine; construct; plasmid; SV-Pewt; flavivirus; infection; prM/M;  
 KW tick-borne encephalitis; TBE; E protein; ss.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2166  
 FT /\*tag= a  
 FT /product= "SV-Pewt"  
 XX  
 EP869184-A2.  
 XX  
 PD 07-OCT-1998.  
 XX  
 PF 06-JUL-1995; 95EP-0890132.  
 XX  
 PR 23-MAY-1995; 95AT-0000871.  
 PR 28-JUL-1994; 94AT-0001352.  
 XX  
 PA (IMMO ) IMMUNO AG.  
 XX  
 PI Allison S, Heinz FX, Kunz C, Mandl CD;  
 XX  
 DR WPI; 1998-508495/44.  
 DR P-PSDB; AAW77407.  
 XX  
 XX Vaccine against flavivirus infection, especially tick-borne  
 PT encephalitis - comprises full-length native protein E and prM/M  
 PT coding sequences  
 XX

PS Example 1; Fig 3A-C; 34pp; German.  
 XX  
 CC This sequence encodes a novel plasmid construct, SV-Pewt which is used  
 CC in a method for the production of a vaccine against flavivirus  
 CC infections. The construct comprises nucleic acid encoding flavivirus  
 CC proteins E and prM/M in full-length native form. The vaccine is used  
 CC especially for vaccination against tick-borne encephalitis (TBE).  
 XX  
 SQ Sequence 2181 BP; 558 A; 450 C; 688 G; 485 T; 0 other;  
 Query Match 79.2%; Score 19.8; DB 19; Length 2181;  
 Best Local Similarity 91.3%; Pred. No. 11;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 TAGGCTGTTTGGAAAGGCTAGCA 23  
 Db 900 TGGACTGTTTGGAAAGGCTAGCA 922  
 RESULT 5  
 AAN81162  
 ID AAN81162 standard; DNA; 2418 BP.  
 XX  
 AC AAN81162;  
 XX  
 DT 26-OCT-1990 (first entry)  
 XX  
 DE Encodes Western subtype of early summer meningoencephalitis (ESME).  
 DE early summer meningoencephalitis virus; live vaccines; ds.  
 KW  
 XX  
 OS Early summer meningoencephalitis virus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 113..460  
 FT /\*tag= a  
 FT /product=protein C  
 FT CDS 461..727  
 FT /\*tag= b  
 FT /product=protein prM  
 FT CDS 728..952  
 FT /\*tag= c  
 FT /product=protein C  
 FT CDS 953..2418  
 FT /\*tag= d  
 FT /product=protein E  
 XX  
 EP284791-A.  
 XX  
 PD 05-DEC-1988.  
 XX  
 PF 29-FEB-1988; 88EP-0103003.  
 XX  
 PR 20-MAR-1987; 87EP-0104114.  
 XX  
 PA (IMMU-) IMMUNO CHEM MED AG.  
 XX  
 PI Heinz FX, Kunz C, Mandl C, Dörner F, Bodemer W;  
 XX  
 DR WPI; 1988-294138/42.  
 DR P-PSDB; AAP80573, AAP82324, AAP82325 & AAP82326.  
 XX  
 XX New DNA and RNA mols encoding proteins of meningoencephalitis virus -  
 PT useful in vaccines, diagnostic agents and detection probes  
 PT  
 PS Disclosure; i p; German.  
 XX  
 CC Encodes all the structural proteins of ESME virus. The invention  
 CC covers fragments of this sequence and analogous RNA molecules.  
 CC Corresponding mRNA sequence given in specification.  
 XX  
 SQ Sequence 2418 BP; 635 A; 507 C; 743 G; 533 T; 0 other;



FT		/tag= k	"nucleotide illegible in the specification"
FT	misc_difference	8460	
FT		/tag= l	"nucleotide illegible in the specification"
FT	misc_difference	8520	
FT		/tag= m	"nucleotide illegible in the specification"
FT	misc_difference	8580	
FT		/tag= n	"nucleotide illegible in the specification"
FT	misc_difference	8760	
FT		/tag= o	"nucleotide illegible in the specification"
XX	DE4426622-CI.		
XX	22-FEB-1996.		
XX	27-JUL-1994;	94DE-4426622.	
XX	27-JUL-1994;	94DE-4426622.	
XX	(IMMO ) IMMUNO AG.		
XX	Heinz FX, Kunz C, Mandl C;		
XX	WPI; 1996-106416/12.		
XX	Complete DNA encoding tick borne encephalitis virus RNA - able to produce infectious viral particles for use in vaccines		
XX	Claim 1; Fig 5; 33pp; German.		
CC	The new complete DNA encoding tick borne encephalitis virus RNA may be used to produce infectious viral particles. These particles may be used in vaccines (live attenuated or inactivated), opt. formulated with an immunomodulator. These virus particles are also useful as seed for viral replication. The new cDNA is also useful itself as a vaccine.		
CC	The sequence is the first complete TBEV sequence known and makes possible targeted alteration of the viral genome (partic. for attenuation or increasing viral yields).		
XX	Sequence 11141 BP; 2820 A; 2455 C; 3526 G; 2326 T; 14 other;		
SQ	Query Match	79.2%; Score 19.8; DB 17; Length 11141;	
	Best Local Similarity	91.3%; Pred. No. 15;	
	Matches	21; Conservative 0; Mismatches 2; Indels 0; Gap	
Oy	1 TAGGCTGTTTGGAAAGGGTAGCA 23		
Dd	1287 TGGACTGTTTGGAAAGGGTAGCA 1309		
RESULT 8			
AAD13425			
ID	AAD13425 standard; cDNA; 5098 BP.		
XX	AAD13425;		
AC			
XX	06-NOV-2001 (first entry)		
DT	Murine Kit/stem cell factor receptor kinase insert region cDNA.		
DE			
XX	Mouse; fertility defect; non-phosphorylated tyrosine; sterility tes		
KW	kinase insert region; Kit/Stem cell factor receptor; Kit/SCF-R;		
KW	Kit; spermatogenesis inhibition; cancer therapy; contraceptive; ss.		
XX			
OS	Mus musculus.		
XX			
Key	Location/Qualifiers		
FH	29..2956		
FT	CDS		

FT	/*tag= a
FT	/product= "Murine Kit/stem cell factor receptor kinase
FT	insert region"
XX	
PN	WO200151672-A2.
XX	
PD	19-JUL-2001.
XX	
XX	08-JAN-2001; 2001WO-US00573.
XX	
PF	11-JAN-2000; 2000US-0175625.
XX	
PR	(SALK ) SALK INST BIOLOGICAL STUDIES.
XX	
PA	Blume-Jensen P, Hunter T;
XX	
PI	WPI; 2001-502588/55.
XX	
DR	P-PSDB; AAE07144.
DR	
XX	Identifying a fertility defect in a male, comprises detecting the
PT	presence of mutation or a non-phosphorylated tyrosine residue in the
PT	kinase insert region of a Kit/Stem cell factor receptor -
XX	
PS	Claim 4; Page 46-47; 54pp; English.
XX	
CC	The present invention relates to a method for identifying a
CC	fertility defect in a male, comprises detecting the presence of a
CC	non-phosphorylated tyrosine residue in the kinase insert region of a
CC	Kit/Stem cell factor receptor (SCF-R) (designated as Kit/SCF-R or Kit),
CC	or a mutation in the nucleic acid or polynucleotide encoding the
CC	Kit/SCF-R. The presence of such mutation or non-phosphorylation indicates
CC	a fertility defect. The assay is useful for identifying or distinguishing
CC	fertility defects in males. The Kit inhibiting agents are suitable as for
CC	incorporation into pharmaceuticals to treat subjects in need of
CC	spermatogenesis inhibition, e.g. for contraception or to preserve
CC	fertility by inhibiting germ cell proliferation during cancer therapy.
CC	The Kit enhancing agents may be used to promote spermatogenesis to
CC	enhance fertility in males suffering from a kit-mediated fertility
CC	defect. The genetically altered mice are useful models for sterility
CC	testing in a mammal. The present sequence is a cDNA encoding murine
CC	kinase insert region of Kit/SCF-R.
XX	
SQ	Sequence 5098 BP; 1309 A; 1224 C; 1235 G; 1330 T; 0 other;
	Query Match            76.8%; Score 19.2; DB 22; Length 5098;
	Best Local Similarity   87.5%; Pred. No. 25;
	Matches   21; Conservative   0; Mismatches   3; Indels   0; Gaps   0;
QY	2 AGGCTGTGGTGAAGGTTAGCACA 25 
Db	3956 AGGCTGTGGTGAAGGTTAGCACA 3979
RESULT 9	
AAX59954	ID
XX	AAX59954 standard; DNA; 2055 BP.
XX	
AC	AAX59954;
XX	
DT	02-AUG-1999 (first entry)
DE	
XX	DNA encoding premembrane and envelope antigenic proteins.
XX	
KW	Premembrane protein; envelope protein; antigen; DNA vaccine;
KW	viral tick-borne encephalitis; cross-protection;
KW	Russian spring summer encephalitis virus;
KW	Central European encephalitis virus; ss.
XX	
OS	Central european encephalitis virus.
XX	
PN	WO9926653-A1.
XX	
PD	03-JUN-1999.



```
FT CDS 39323..43459
FT /tag= v
FT /product= "major capsid protein"
FT /label= RRV_ORF25
FT /note= "has similarity to KSHV ORF25"
FT 43491..44408
FT /tag= w
FT /product= "capsid protein"
FT /label= RRV_ORF26
FT /note= "has similarity to KSHV ORF26"
FT 44433..45242
FT /tag= x
FT /label= RRV_ORF27
FT /note= "has similarity to KSHV ORF27"
FT 45408..45683
FT /tag= y
FT /label= RRV_ORF28
FT /note= "has similarity to KSHV ORF28"
FT complement (45733..46779)
FT /tag= z
FT /label= RRV_ORF29b
FT /note= "has similarity to KSHV ORF29b"
FT 46905..47135
FT /tag= aa
FT /label= RRV_ORF30
FT /note= "has similarity to KSHV ORF30"
FT 47093..47746
FT /tag= ab
FT /label= RRV_ORF31
FT /note= "has similarity to KSHV ORF31"
FT 47683..49077
FT /tag= ac
FT /label= RRV_ORF32
FT /note= "has similarity to KSHV ORF32"
FT 49049..50059
FT /tag= ad
FT /label= RRV_ORF33
FT /note= "has similarity to KSHV ORF33"
FT complement (49977..50960)
FT /tag= ae
FT /label= RRV_ORF29a
FT /note= "has similarity to KSHV ORF29a"
FT 50959..51942
FT /tag= af
FT /label= RRV_ORF34
FT /note= "has similarity to KSHV ORF34"
FT 51923..52372
FT /tag= ag
FT /label= RRV_ORF35
FT /note= "has similarity to KSHV ORF35"
FT 52278..53585
FT /tag= ah
FT /product= "kinase"
FT /label= RRV_ORF36
FT /note= "has similarity to KSHV ORF36"
FT 53566..55008
FT /tag= ai
FT /product= "alkaline exonuclease"
FT /label= RRV_ORF37
FT /note= "has similarity to KSHV ORF37"
FT 54963..55172
FT /tag= aj
FT /label= RRV_ORF38
FT /note= "has similarity to KSHV ORF38"
FT complement (55255..56391)
FT /tag= ak
FT /product= "glycoprotein M"
FT /label= RRV_ORF39
FT /note= "has similarity to KSHV ORF39"
FT 56526..57932
FT /tag= al
FT /product= "helicase/primase"
FT /label= RRV_ORF40

/Note= "has similarity to KSHV ORF40"
57917..58528
/tag= am
/product= "helicase/primase"
/label= RRV_ORF41
/Note= "has similarity to KSHV ORF41"
complement (58525..59343)
/tag= an
/label= RRV_ORF42
/Note= "has similarity to KSHV ORF42"
complement (59297..61027)
/tag= ao
/product= "capsid protein"
/label= RRV_ORF43
/Note= "has similarity to KSHV ORF43"

Query Match 72.8%; Score 18.2; DB 24; Length 128139;
Best Local Similarity 87.0%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GCCTGTTTGGAAAGGTAGCACCA 25
||||| ||||| ||||| |||||
Db 27481 GGCTGGTGGAAACGGTAGCACCA 27503

RESULT 11
AAC64754
ID AAC64754 standard; DNA; 133719 BP.
XX
AC AAC64754;
XX
DT 28-FEB-2001 (first entry)
XX
DE Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
KW Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
splenomegaly; hypergammaglobulinaemia; autoimmune haemolytic anaemia;
ds.
XX
OS Macaca mulatta rhadinovirus 17577.
XX
PN WO200028040-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26260.
XX
PR 06-NOV-1998; 98US-0107507.
PR 20-NOV-1998; 98US-0109409.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Wong SW, Axthelm MK, Searles RP;
XX
WPI; 2000-376552/32.
XX
PT New rhesus rhadino virus for producing non-human primate model useful
PT for testing potential treatments and efficacy of the candidate vaccine
PT for conditions associated with RRV infection -
XX
PS Claim 2; Page 83-122; 141pp; English.
XX
CC The present invention describes a novel rhesus macaque rhadinovirus
called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
encoded by the genome sequence. The present invention also specifically
claims the individual open reading frame (ORF) nucleotide sequences from
the genome which encode the individual proteins, but these sequences are
not given. A non-human animal infected with RRV can be used for testing
```

CC the efficacy of drug in the treatment of condition associated with  
CC infection with RRV such as Kaposi's sarcoma, lymphoproliferative  
CC disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,  
CC hypergammaglobulinemia or autoimmune haemolytic anaemia, by  
CC administering the drug to a immuno-compromised non-human primate  
CC preferably Rhesus macaque monkey obtained by as a result of infection  
CC by Simian Immunodeficiency Virus (SIV). RRV is useful for producing  
CC non-human primate model for testing potential treatments for conditions  
CC associated with RRV infection. It is also useful for testing the  
CC efficacy of the candidate vaccine against RRV infection or conditions  
CC associated with its infection by administering the vaccine to the  
CC subject capable of infection with RRV, inoculating the subject with RRV  
CC and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205  
CC to AAB53213 represent sequence used in the exemplification of the  
CC present invention.  
XX  
SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 T; 0 other;  
Query Match 72.8%; Score 18.2; DB 21; Length 133719;  
Best Local Similarity 87.0%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 3 GCCTCTTGGAAAGGAGGTAGCACA 25  
||||| | ||||| ||||| |||||  
Db 33061 GGCTGGTGGAAACGAGGTAGCACA 33083  
RESULT 12  
ABN95473  
ID ABN95473 standard; DNA; 372 BP.  
XX AC ABN95473;  
XX DT 13-AUG-2002 (first entry)  
XX DE Gene #1971 used to diagnose liver cancer.  
XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX  
OS Homo sapiens.  
XX PN WO200229103-A2.  
XX PD 11-APR-2002.  
XX PF 02-OCT-2001; 2001WO-US30589.  
XX PR 02-OCT-2000; 2000US-237054P.  
XX PA (GENE-) GENE LOGIC INC.  
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX WPI; 2002-426119/45.  
XX DR  
XX PT Diagnosing and detecting the progression of liver cancer,  
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
XX PT involves detecting the level of expression of two or more genes in a  
XX PT liver tissue sample -  
XX PS Claim 1; SEQ ID NO 1971; 298pp; English.  
XX CC The invention relates to a novel method for diagnosing and detecting the  
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
XX CC tumour in a patient, and differentiating metastatic liver cancer from  
XX CC hepatocellular carcinoma in a patient, involving detecting the level of  
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a  
XX CC tissue sample. The method of the invention has hepatotropic, and  
XX CC cytostatic activity. The method is useful for diagnosing and detecting  
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
XX CC liver carcinoma in a patient. The method is useful for identifying

CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 372 BP; 119 A; 55 C; 67 G; 118 T; 13 other;  
Query Match 72.0%; Score 18; DB 24; Length 372;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TAGGCTGTTTGGAAAGGG 18  
||||| | ||||| ||||| |||||  
Db 279 TAGGCTGTTTGGAAAGGG 296  
RESULT 13  
ABL65082  
ID ABL65082 standard; DNA; 372 BP.  
XX AC ABL65082;  
XX DT 15-MAY-2002 (first entry)  
XX DE Lung cancer related gene sequence SEQ ID NO:3419.  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX OS Homo sapiens.  
XX PN WO200194629-A2.  
XX PD 13-DEC-2001.  
XX PF 30-MAY-2001; 2001WO-US10838.  
XX PR 05-JUN-2000; 2000US-209473P.  
XX PR 05-JUN-2000; 2000US-209531P.  
XX PR 18-SEP-2000; 2000US-233133P.  
XX PR 18-SEP-2000; 2000US-233617P.  
XX PR 20-SEP-2000; 2000US-234009P.  
XX PR 20-SEP-2000; 2000US-234034P.  
XX PR 20-SEP-2000; 2000US-234052P.  
XX PR 22-SEP-2000; 2000US-234059P.  
XX PR 22-SEP-2000; 2000US-234567P.  
XX PR 25-SEP-2000; 2000US-234923P.  
XX PR 25-SEP-2000; 2000US-234924P.  
XX PR 25-SEP-2000; 2000US-235077P.  
XX PR 25-SEP-2000; 2000US-235082P.  
XX PR 25-SEP-2000; 2000US-235134P.  
XX PR 25-SEP-2000; 2000US-235280P.  
XX PR 26-SEP-2000; 2000US-235637P.  
XX PR 26-SEP-2000; 2000US-235638P.  
XX PR 27-SEP-2000; 2000US-235711P.  
XX PR 27-SEP-2000; 2000US-235720P.  
XX PR 27-SEP-2000; 2000US-235840P.  
XX PR 27-SEP-2000; 2000US-235863P.  
XX PR 28-SEP-2000; 2000US-236028P.  
XX PR 28-SEP-2000; 2000US-236032P.  
XX PR 28-SEP-2000; 2000US-236033P.  
XX PR 28-SEP-2000; 2000US-236034P.  
XX PR 28-SEP-2000; 2000US-236109P.  
XX PR 28-SEP-2000; 2000US-236111P.  
XX PR 29-SEP-2000; 2000US-236842P.  
XX PR 29-SEP-2000; 2000US-236891P.  
XX PR 02-OCT-2000; 2000US-237172P.  
XX PR 02-OCT-2000; 2000US-237173P.  
XX PR 02-OCT-2000; 2000US-237278P.





CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 372 BP; 119 A; 55 C; 67 G; 118 T; 13 other;

Query Match 72.0%; Score 18; DB 24; Length 372;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTTGGAAAGGG 18  
IIIIIIIIIIIIIIIIIIII  
Db 279 TAGGCTGTTTGGAAAGGG 296

## RESULT 15

ABL66942  
ID ABL66942 standard; DNA; 372 BP.

XX  
AC ABL66942;

XX 15-MAY-2002 (first entry)

DT Lung cancer related gene sequence SEQ ID NO:5279.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
gene; ds.

XX Homo sapiens.

OS  
XX WO200194629-A2.

PN 13-DEC-2001.

PD 30-MAY-2001; 2001WO-US10838.

XX 05-JUN-2000; 2000US-209473P.

PR 05-JUN-2000; 2000US-209531P.

PR 18-SEP-2000; 2000US-233133P.

PR 18-SEP-2000; 2000US-233617P.

PR 20-SEP-2000; 2000US-234009P.

PR 20-SEP-2000; 2000US-234034P.

PR 22-SEP-2000; 2000US-234052P.

PR 22-SEP-2000; 2000US-234509P.

PR 22-SEP-2000; 2000US-234567P.

PR 23-SEP-2000; 2000US-234923P.

PR 23-SEP-2000; 2000US-234924P.

PR 25-SEP-2000; 2000US-235077P.

PR 25-SEP-2000; 2000US-235082P.

PR 25-SEP-2000; 2000US-235134P.

PR 25-SEP-2000; 2000US-235280P.

PR 26-SEP-2000; 2000US-235637P.

PR 26-SEP-2000; 2000US-235638P.

PR 27-SEP-2000; 2000US-235711P.

PR 27-SEP-2000; 2000US-235720P.

PR 27-SEP-2000; 2000US-235840P.

PR 27-SEP-2000; 2000US-235863P.

PR 28-SEP-2000; 2000US-236028P.

PR 28-SEP-2000; 2000US-236032P.

PR 28-SEP-2000; 2000US-236033P.

PR 28-SEP-2000; 2000US-236034P.

PR 28-SEP-2000; 2000US-236109P.

PR 28-SEP-2000; 2000US-236111P.

PR 29-SEP-2000; 2000US-236842P.

PR 29-SEP-2000; 2000US-236891P.

PR 02-OCT-2000; 2000US-237172P.

PR 02-OCT-2000; 2000US-237173P.

PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.

PR 02-OCT-2000; 2000US-237295P.

PR 02-OCT-2000; 2000US-237316P.

PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
XX (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX Claim 1; SEQ ID 5279; 44pp; English.

XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.

XX Sequence 372 BP; 119 A; 55 C; 67 G; 118 T; 13 other;

Query Match 72.0%; Score 18; DB 24; Length 372;

Best Local Similarity 100.0%; Pred. No. 59;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGCTGTTTGGAAAGGG 18

IIIIIIIIIIIIIIIIIIII

Db 279 TAGGCTGTTTGGAAAGGG 296

Search completed: January 18, 2003, 23:46:12

Job time : 256.5 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2003, 22:00:24 ; Search time 212.5 seconds  
(without alignments)  
264.941 Million cell updates/sec

Title: US-09-674-593-8  
Perfect score: 25  
Sequence: 1 taaatgggtggcggtggggagac 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	25	100.0	25	AAZ36647	PCR primer VDE119
2	25	100.0	1382	AAZ36643	Human tumour rejec
3	20	80.0	339	AAI88448	Human polynucleoti
4	19.8	79.2	12141	AAI62854	Mouse PEPCK-cytoso
5	19.2	76.8	1000	AAI57997	Arachidonic acid m
6	19.2	76.8	1000	AAI57998	Arachidonic acid m
7	19.2	76.8	1000	AAI57999	Arachidonic acid m
8	19.2	76.8	1000	AAI58000	Arachidonic acid m
9	19.2	76.8	9594	ABL10394	Drosophila melanog

C	10	19.2	76.8	20674	21	AAC58017	Arachidonic acid m
C	11	18.6	74.4	335	23	ABV25168	Human prostate exp
C	12	18.6	74.4	469	23	ABV37980	Human prostate exp
C	13	18.6	74.4	1087	22	AAH68547	Human protein HP10
C	14	18.6	74.4	1089	21	AAA40516	Murine adult splee
C	15	18.6	74.4	1098	22	AAI60823	Human polynucleoti
C	16	18.6	74.4	1243	22	AAI59037	Human polynucleoti
C	17	18.6	74.4	2605	19	AAV04699	Homo sapiens 20q13
C	18	18.2	72.8	347	24	ABN61824	Human cancer relat
C	19	18.2	72.8	823	22	AAI06126	Human reproductive
C	20	18.2	72.8	823	22	AAI06127	Human reproductive
C	21	18.2	72.8	823	23	ABL98691	Human testicular a
C	22	18.2	72.8	823	23	ABL98692	Human testicular a
C	23	18.2	72.8	853	22	AAI525893	Human testicular a
C	24	18.2	72.8	1738	22	AAI04956	Human cDNA encodin
C	25	18.2	72.8	1738	22	AAI04957	Human reproductive
C	26	18.2	72.8	1738	23	ABL97850	Human testicular a
C	27	18.2	72.8	1738	23	ABL97851	Human testicular a
C	28	18.2	72.8	9551	22	AAI39928	Genomic sequence #
C	29	18.2	72.8	9551	22	AAI90372	Human digestive sy
C	30	18.2	72.8	15558	22	AAI39926	Genomic sequence #
C	31	18.2	72.8	15558	22	AAI90370	Human digestive sy
C	32	18.2	72.8	22893	24	ABK12810	Human tumour suppr
C	33	18.2	72.8	30625	24	ABK12808	Human tumour suppr
C	34	18	72.0	211	21	AAI30165	Human secreted pro
C	35	17.8	71.2	531	24	ABQ41640	Oligonucleotide fo
C	36	17.8	71.2	531	24	ABQ41641	Oligonucleotide fo
C	37	17.6	70.4	252	21	AAI08438	Human secreted pro
C	38	17.6	70.4	573	24	ABQ48008	Oligonucleotide fo
C	39	17.6	70.4	573	24	ABQ48009	Oligonucleotide fo
C	40	17.6	70.4	610	24	ABQ45844	Oligonucleotide fo
C	41	17.6	70.4	610	24	ABQ45845	Oligonucleotide fo
C	42	17.6	70.4	1154	24	ABL39899	Human NS cDNA sequ
C	43	17.6	70.4	1613	23	AAI92512	DNA encoding novel
C	44	17.6	70.4	1682	23	AAI583590	DNA encoding novel
C	45	17.6	70.4	2747	24	ABL34397	Human immune syste

ALIGNMENTS

RESULT 1  
AAZ36647  
ID AAZ36647 standard; cDNA; 25 BP.  
XX AAZ36647;  
AC AAZ36647;  
XX  
XX  
DT 22-FEB-2000 (first entry)  
DE PCR primer VDE119 used to amplify RUR-1 antisense cDNA sequence.  
XX  
XX  
XX Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;  
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;  
KW Leukaemia; PCR primer; ss.  
XX  
XX OS Synthetic.  
OS Homo sapiens.  
XX  
XX PN WO9558546-A1.  
XX  
XX PD 18-NOV-1999.  
XX  
XX PF 13-MAY-1999; 99WO-0510424.  
XX  
XX PR 13-MAY-1998; 98US-0085318.  
XX  
XX PA (LUDW-) LUDWIG INST CANCER RES.  
XX  
XX PI Van Den Eynde B, Boon-Falleur T;  
XX  
XX DR WPI; 2000-053076/04.  
XX  
XX PT New isolated tumour rejection antigen RUR-1 nucleic acids, used for,

PT e.g. treatment of cancers -

PS Example 1; Page 44; 75pp; English.

XX

CC PCR primers AA226647-48 were used to amplify the antisense cDNA sequence

CC of human tumour rejection antigen RUR-1. The RUR-1 antisense sequence

CC is the antisense strand of a ubiquitously expressed gene. The antisense

CC strand codes for a polypeptide which is preferentially expressed in

CC tumour samples and tumour-derived cells lines. The polypeptide is

CC unrelated to any TRAP protein. The sequence was isolated from a renal

CC cell carcinoma line LB9211-RCC. The RUR-1 nucleic acids and polypeptides

CC can be used for diagnosis, prognosis or treatment of a disorder

CC characterized by the expression of a RUR-1 antisense cDNA molecule or

CC an expression product, such as cancers, e.g. renal cell carcinoma,

CC colorectal carcinoma, melanoma, sarcoma or leukaemia.

XX

SQ Sequence 25 BP; 5 A; 2 C; 14 G; 4 T; 0 other;

Query Match 100.0%; Score 25; DB 21; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.46;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAGAC 25

|||||

Db 1 TAAATGGGTGGCGGTGGGGGAGAC 25

RESULT 2

AAZ36643

ID AAZ36643 standard; cDNA; 1382 BP.

XX

AC AAZ36643;

XX

DT 22-FEB-2000 (first entry)

XX

DE Human tumour rejection antigen RUR-1 antisense cDNA sequence.

XX

KW Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;

KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma;

KW leukaemia; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 738..992

FT /\*tag= a

FT primer\_bind 523..547

FT /\*tag= b

FT /\*note= "binding site for primer VDE119 (see AAZ36647)"

FT primer\_bind complement (1280..1305)

FT /\*tag= c

FT /\*note= "binding site for primer VDE120 (see AAZ36648)"

XX

PN WO9958546-A1.

XX

PD 18-NOV-1999.

XX

PF 13-MAY-1999; 99WO-US10424.

XX

PR 13-MAY-1998; 98US-0085318.

XX

PA (LUDW-) LUDWIG INST CANCER RES.

XX

PI Van Den Eynde B, Boon-Falleur T;

XX

DR WPI; 2000-053076/04.

DR P-PSDB; AAY53809.

XX

PT New isolated tumour rejection antigen RUR-1 nucleic acids, used for,

PT e.g. treatment of cancers -

XX

PS Claim 4; Fig 5; 75pp; English.

XX

CC The present sequence represents the antisense cDNA sequence of human

CC tumour rejection antigen RUR-1. The present sequence is the antisense

CC strand of a ubiquitously expressed gene. The antisense strand codes for

CC a polypeptide which is preferentially expressed in tumour samples and

CC tumour-derived cells lines. The polypeptide is unrelated to any TRAP

CC protein. The sequence was isolated from a renal cell carcinoma line

CC LB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for

CC diagnosis, prognosis or treatment of a disorder characterized by the

CC expression of a RUR-1 antisense cDNA molecule or an expression product,

CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,

CC melanoma, sarcoma or leukaemia.

XX

SQ Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 other;

Query Match 100.0%; Score 25; DB 21; Length 1382;

Best Local Similarity 100.0%; Pred. No. 0.54;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAGAC 25

|||||

Db 523 TAAATGGGTGGCGGTGGGGGAGAC 547

RESULT 3

AAI88448

ID AAI88448 standard; cDNA; 339 BP.

XX

AC AAI88448;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 8508.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation; ss.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US04927.

XX

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

DR WPI; 2001-514838/56.

DR P-PSDB; AAO08517.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX

PS Claim 1; SEQ ID NO 8508; 1399pp + Sequence Listing; English.

XX

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and

CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 339 BP; 84 A; 72 C; 135 G; 48 T; 0 other;

Query Match 80.0%; Score 20; DB 22; Length 339;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 GGGTGGCGGTGGGGGAGAC 25  
|||||

Db 287 GGGTGGCGGTGGGGGAGAC 306

## RESULT 4

AAF62854/C  
ID AAF62854 standard; DNA; 12141 BP.

XX AC AAF62854;

XX 08-MAY-2001 (first entry)

XX Mouse PEPCCK-cytosolic nucleotide sequence.

XX Mouse; antiinflammatory; cytostatic; antisense gene therapy;  
KW phosphoenol pyruvate carboxykinase-cytosolic; PEPCCK-cytosolic;  
KW infection; inflammation; tumour formation; ds.

XX Mus musculus.

XX US6187545-B1.

XX 13-FEB-2001.

XX 21-JAN-2000; 2000US-0488671.

XX 21-JAN-2000; 2000US-0488671.

XX (ISIS-) ISIS PHARM INC.

PI McKay R, Butler MM, Wyatt J, Cowser LM;

XX WPI; 2001-190979/19.

XX Antisense compound capable of modulating the expression of phosphoenol  
PT pyruvate carboxykinase-cytosolic, useful for preventing or delaying  
PT infection, inflammation or tumor formation -

PS Example 13; Column 55-70; 64pp; English.

XX The present sequence was used in an example illustrating an invention  
CC relating to antisense compounds, compositions and methods for modulating  
CC the expression of phosphoenol pyruvate carboxykinase-cytosolic  
CC (PEPCCK-cytosolic). Antisense compounds of up to 30 nucleobases in length  
CC that are capable of inhibiting the expression of PEPCCK-cytosolic are  
CC provided. The antisense compounds are useful for inhibiting the  
CC expression of PEPCCK-cytosolic in cells or tissues. They are commonly  
CC used as research reagents and in diagnostics, e.g. to elucidate the  
CC function of particular genes. They are also useful for distinguishing  
CC between functions of various members of a biological pathway and for  
CC research use. The antisense compounds are also useful prophylactically,  
CC e.g. to prevent or delay infection, inflammation or tumour formation.

XX Sequence 12141 BP; 3122 A; 2924 C; 3087 G; 3008 T; 0 other;

Query Match 79.2%; Score 19.8; DB 22; Length 12141;  
Best Local Similarity 91.3%; Pred. No. 68;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAAATGGGTGGCGGTGGGGGAG 23

|||||

Db 8237 TAAATGGGAGGAGGTGGGGAG 8215

## RESULT 5

AAC57997  
ID AAC57997 standard; DNA; 1000 BP.

XX AC AAC57997;

XX 25-JAN-2001 (first entry)

XX Arachidonic acid metabolism related genomic biallelic marker #631.

XX Human; biallelic marker; arachidonic acid metabolism; genotyping;  
KW detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;  
KW single nucleotide polymorphism; hybridisation assay; sequencing assay;  
KW specific amplification assay; identification; ERBM; 12-LO-RBM;  
KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.

XX Homo sapiens.

XX WO200047771-A2.

XX 17-AUG-2000.

XX 11-FEB-2000; 2000WO-IB00184.

XX 12-FEB-1999; 99US-0119917.

XX 23-MAR-1999; 99US-0275267.

XX 07-MAY-1999; 99US-0133200.

XX (GEST ) GENSET.

XX Blumenfeld M, Bougueleret L, Chumakov I;

XX WPI; 2000-571881/53.

XX Novel biallelic markers useful for detecting conditions and genotypes  
XX associated with arachidonic acid metabolism -

XX Claim 13; Page 754-755; 802pp; English.

XX The present invention describes polynucleotides including biallelic  
CC markers derived from genes involved in arachidonic acid metabolism and  
CC from genomic regions flanking those genes. Methods from the present  
CC invention may be used to select individuals for clinical trials and  
CC predict responses to treatment with drugs. The polynucleotides may be  
CC used in hybridisation assays, sequencing assays and specific  
CC amplification assays for identifying an eicosanoid-related biallelic  
CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a  
CC segment of nucleotides containing an ERBM. The polynucleotides are  
CC useful in diagnostic kits. The markers may be used to detect conditions  
CC and genotypes associated with arachidonic acid metabolism. AAC57367 to  
CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the  
CC exemplification of the present invention.

CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as  
CC SNPs) in the polynucleotide sequences from the present invention have  
CC been given as their corresponding degenerate bases e.g. a polymorphic  
CC base of C or T has been given as Y.

XX Sequence 1000 BP; 210 A; 291 C; 291 G; 207 T; 1 other;

Query Match 76.8%; Score 19.2; DB 21; Length 1000;  
Best Local Similarity 87.5%; Pred. No. 11e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATGGGTGGCGGTGGGGGAGAC 25

|||||

Db 204 AAATGGGTGGCGGTGGGGGAGAC 227

## RESULT 6

AAC57998

ID AAC57998 standard; DNA; 1000 BP.



KW specific amplification assay; identification; ERM; 12-LO-RBM;  
 KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.

OS Homo sapiens.

PN WO200047771-A2.

XX 17-AUG-2000.

XX 11-FEB-2000; 2000WO-IB00184.

XX 12-FEB-1999; 99US-0119917.

PR 23-MAR-1999; 99US-0275267.

XX 07-MAY-1999; 99US-0133200.

PA (GEST ) GENSET.

PI Blumenfeld M, Bougueleret L, Chumakov I;

XX WPI; 2000-571881/53.

XX Novel biallelic markers useful for detecting conditions and genotypes  
 PT associated with arachidonic acid metabolism -  
 XX Claim 13; Page 757; 802pp; English.

XX The present invention describes polynucleotides including biallelic  
 CC markers derived from genes involved in arachidonic acid metabolism and  
 CC from genomic regions flanking those genes. Methods from the present  
 CC invention may be used to select individuals for clinical trials and  
 CC predict responses to treatment with drugs. The polynucleotides may be  
 CC used in hybridisation assays, sequencing assays and specific  
 CC amplification assays for identifying an eicosanoid-related biallelic  
 CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a  
 CC segment of nucleotides containing an ERM. The polynucleotides are  
 CC useful in diagnostic kits. The markers may be used to detect conditions  
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to  
 CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the  
 CC exemplification of the present invention.  
 CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as  
 CC SNPs) in the polynucleotide sequences from the present invention have  
 CC been given as their corresponding degenerate bases e.g. a polymorphic  
 CC base of C or T has been given as Y.

SQ Sequence 1000 BP; 222 A; 299 C; 265 G; 213 T; 1 other;

Query Match 76.8%; Score 19.2; DB 21; Length 1000;

Best Local Similarity 87.5%; Pred. NO. 1.1e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATGGTGGCGGTGGGGGAGAC 25

||||| ||||| ||||| |||||

Db 10 AAATGGTGGCGGTGGGGGAGAC 33

RESULT 9

ABL10394

ID ABL10394 standard; cDNA; 9594 BP.

XX ABL10394;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25664.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB66291.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX Claim 1; SEQ ID NO 25664; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB872072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 9594 BP; 2719 A; 2040 C; 2045 G; 2790 T; 0 other;

Query Match 76.8%; Score 19.2; DB 23; Length 9594;

Best Local Similarity 87.5%; Pred. NO. 1.2e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATGGTGGCGGTGGGGGAGAC 25

||||| ||||| ||||| |||||

Db 3005 AAATGGTGGCGGTGGGGGAGAC 3028

RESULT 10

AAC58017

ID AAC58017 standard; DNA; 20674 BP.

XX AAC58017;

XX 25-JAN-2001 (first entry)

XX Arachidonic acid metabolism related genomic biallelic marker #651.

XX Human; biallelic marker; arachidonic acid metabolism; genotyping;  
 DE detection; hybridisation; phenotype; haplotype; SNP; polymorphic base;  
 KW single nucleotide polymorphism; hybridisation assay; sequencing assay;  
 KW specific amplification assay; identification; ERM; 12-LO-RBM;  
 KW eicosanoid-related biallelic marker; 12-LO-related biallelic marker; ds.

XX Homo sapiens.

XX WO200047771-A2.

XX 17-AUG-2000.

XX 11-FEB-2000; 2000WO-IB00184.

XX 12-FEB-1999; 99US-0119917.

PR 23-MAR-1999; 99US-0275267.

PR 07-MAY-1999; 99US-0133200.

XX (GEST ) GENSET.

XX Blumenfeld M, Bougueleret L, Chumakov I;

XX WPI; 2000-571881/53.  
 XX Novel biallelic markers useful for detecting conditions and genotypes  
 PT associated with arachidonic acid metabolism -  
 XX Claim 67; Page 790-796; 802pp; English.  
 XX The present invention describes polynucleotides including biallelic  
 CC markers derived from genes involved in arachidonic acid metabolism and  
 CC from genomic regions flanking those genes. Methods from the present  
 CC invention may be used to select individuals for clinical trials and  
 CC predict responses to treatment with drugs. The polynucleotides may be  
 CC used in hybridisation assays, sequencing assays and specific  
 CC amplification assays for identifying an eicosanoid-related biallelic  
 CC marker (ERBM) or 12-LO-related biallelic marker, and for amplifying a  
 CC segment of nucleotides containing an ERBM. The polynucleotides are  
 CC useful in diagnostic kits. The markers may be used to detect conditions  
 CC and genotypes associated with arachidonic acid metabolism. AAC57367 to  
 CC AAC58018 and AAB24019 and AAB24020 represent sequences used in the  
 CC exemplification of the present invention.  
 CC N.B. Polymorphic bases (single nucleotide polymorphisms also known as  
 CC SNPs) in the polynucleotide sequences from the present invention have  
 CC been given as their corresponding degenerate bases e.g. a polymorphic  
 CC base of C or T has been given as Y.  
 XX Sequence 20674 BP; 5400 A; 5170 C; 4995 G; 5062 T; 47 other;  
 SQ  
 Query Match 76.8%; Score 19.2; DB 21; Length 20674;  
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 2 AAATGGTGGCGGTGGGGGAGAC 25  
 DB 13044 AAATGGTGGCGGGGGGGGAGAC 13067  
 RESULT 11  
 ABV25168/c  
 ID ABV25168 standard; cDNA; 335 BP.  
 XX AC ABV25168;  
 XX 16-SEP-2002 (first entry)  
 XX Human prostate expression marker cDNA 25159.  
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX Homo sapiens.  
 XX WO200160860-A2.  
 XX 23-AUG-2001.  
 XX 20-FEB-2001; 2001WO-US05171.  
 XX 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PA Schlegel R, Endege WO, Monahan JE;  
 PI WPI; 2001-662795/76.  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT

PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX Claim 1; Page 4916; 11750pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX Sequence 335 BP; 85 A; 82 C; 76 G; 82 T; 10 other;  
 SQ  
 Query Match 74.4%; Score 18.6; DB 23; Length 335;  
 Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 TAAATGGTGGCGGTGGGGGAGAC 25  
 DB 73 TAAAGGGTGGGAGTGGAGGAAAC 49  
 RESULT 12  
 ABV37980  
 ID ABV37980 standard; cDNA; 469 BP.  
 XX AC ABV37980;  
 XX 16-SEP-2002 (first entry)  
 XX Human prostate expression marker cDNA 37971.  
 XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX Homo sapiens.  
 XX WO200160860-A2.  
 XX 23-AUG-2001.  
 XX 20-FEB-2001; 2001WO-US05171.  
 XX 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PA Schlegel R, Endege WO, Monahan JE;  
 PI WPI; 2001-662795/76.  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX Claim 1; Page 7763; 11750pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the



CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 469 BP; 104 A; 108 C; 120 G; 137 T; 0 other;

Query Match 74.4%; Score 18.6; DB 23; Length 469;  
Best Local Similarity 84.0%; Pred. No. 1.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TAAATGGTGGCGGTGGGGAGAC 25  
||||| ||||| ||||| ||||| |||||  
Db 337 TAAAGGGTGGAGGTGGAGAAAC 421

RESULT 13  
AAH68547/c  
ID AAH68547 standard; cDNA; 1087 BP.  
XX  
AC AAH68547;  
XX  
DT 13-SEP-2001 (first entry)  
XX  
DE Human protein HP10149 coding sequence.

XX  
KW Human; gene therapy; tumour; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142302-A1.  
XX  
PD 14-JUN-2001.

XX  
PF 06-DEC-2000; 2000WO-JP08631.

XX  
PR 06-DEC-1999; 99JP-0346863.

XX  
PR 06-DEC-1999; 99JP-0346864.

XX  
PR 08-FEB-2000; 2000JP-0031052.

XX  
PR 10-FEB-2000; 2000JP-0034090.

XX  
PR 10-FEB-2000; 2000JP-0034091.

XX  
PR 14-FEB-2000; 2000JP-0035829.

XX  
PR 14-FEB-2000; 2000JP-0035899.

XX  
PR 14-MAR-2000; 2000JP-0071161.

XX  
PR 30-MAY-2000; 2000JP-0160851.

XX  
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX  
XX  
XX Kato S, Eguchi C, Sasaki M;  
XX  
XX WPI; 2001-381646/40.  
XX  
XX P-PSDB; AAG93262.

XX  
PT Human protein originated from tumor cell line, applicable as drug,  
PT reagent for studying intracellular protein networks and protein source  
PT for drug screening, also encoded cDNA for gene diagnosis and gene  
PT therapy -  
XX

XX  
PS Claim 3; Pages 169-171; 471pp; Japanese.

XX  
XX The present sequence is a human protein coding sequence. The human  
XX protein, preferably originated from tumour cell line, is applicable as a  
XX drug, a reagent for studying intracellular protein networks and a protein  
XX source for screening proteins for binding low molecular weight drugs. The

CC human protein coding sequence is useful for gene diagnosis and gene  
CC therapy, expression vectors and transformant cells for detection of  
CC ligands and receptors.

SQ Sequence 1087 BP; 246 A; 304 C; 288 G; 249 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 1087;  
Best Local Similarity 84.0%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TAAATGGTGGCGGTGGGGAGAC 25  
||||| ||||| ||||| ||||| |||||  
Db 921 TAAAGGGTGGAGGAGGAGGAGAC 897

RESULT 14  
AAA40516/c  
ID AAA40516 standard; cDNA; 1089 BP.

XX  
AC AAA40516;

XX  
DT 16-NOV-2000 (first entry)

XX  
DE Murine adult spleen cDNA fragment AE402\_11.

XX  
KW Secreted protein; cytostatic; immunostimulatory; antimicrobial;  
KW antiviral; immunosuppressive; antinflammatory; vulnerrary; cytokine;  
KW cell proliferation; differentiation; regulator; treatment; tumor;  
KW autoimmune disease; inflammatory disorder; wound; microbial infection;  
KW viral disease; graft versus host reaction suppression; ss.

XX  
OS Mus sp..

XX  
PN WO200037630-A1.

XX  
PD 29-JUN-2000.

XX  
PF 22-DEC-1999; 99WO-US31005.

XX  
PR 23-DEC-1998; 98US-0220876.

XX  
XX (GEMY ) GENETICS INST INC.

XX  
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Bowman MR;

XX  
DR WPI; 2000-442661/38.

XX  
DR P-PSDB; AAB10244.

XX  
PT Secreted human proteins AS296-11 and AS34-11, useful for treating  
PT tumors, autoimmune diseases, inflammatory disorders, wounds, microbial  
PT infections and viral diseases -

XX  
PS Disclosure; Page 225-226; 293pp; English.

XX  
XX This invention describes novel secreted human proteins (I) which have  
XX cytostatic, immunostimulatory, antimicrobial, antiviral,  
XX immunosuppressive, antinflammatory and vulnerrary activity and which act  
XX as cytokine, cell proliferation or differentiation regulators. (I)  
XX is useful for treating tumors, autoimmune diseases, inflammatory  
XX disorders, wounds, microbial infections and viral diseases. (I) is also  
XX useful for suppressing graft versus host reaction. AAA40490-A40580  
XX represent cDNA fragments that encode the secreted proteins  
XX AAB10226-B10288 described in the method of the invention.

SQ Sequence 1089 BP; 263 A; 299 C; 284 G; 243 T; 0 other;

Query Match 74.4%; Score 18.6; DB 21; Length 1089;  
Best Local Similarity 84.0%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TAAATGGTGGCGGTGGGGAGAC 25  
||||| ||||| ||||| ||||| |||||

Db 905 TAAAGGGTGGGAGGAGGAGGAGAC 881

RESULT 15

AAI60823  
ID AAI60823 standard; cDNA; 1098 BP.

XX AC AAI60823;

XX DT 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 4812.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX P-PSDB; AAM41667.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 4812; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX Sequence 1098 BP; 245 A; 287 C; 308 G; 258 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 1098;  
Best Local Similarity 84.0%; Pred. No. 1.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 TAAATGGGTGGCGGTGGGGGAGAC 25  
|||||  
Db 179 TAAAGGGTGGGAGGAGGAGGAGAC 203  
|||||

Search completed: January 18, 2003, 23:45:28  
Job time : 219.5 secs